

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 79.2607 Seconds  
(without alignments)  
388.502 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028

Sequence: 1 TIVSFHQNISNNKTERSTA.....LPDEKEFTCAQSGGQQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

```

1: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/Geneseq/Geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1028	100.0	1426	23	Human legless homo
2	1028	100.0	1435	22	Human BCL9 homolog
3	277	26.9	140	22	Human polypeptide
4	232.5	22.6	320	23	Mouse beta-catenin
5	232.5	22.6	1494	23	Mouse beta-catenin
6	167.5	16.3	707	22	Human polypeptide
7	162.5	15.8	574	22	WASP homolog prote
8	162.5	15.8	574	22	Amino acid sequenc
9	159	15.5	406	22	Novel human diagno

10	158.5	15.4	256	22	ABU52945	Human mammary carc
11	158.5	15.4	270	22	ABU52939	Human mammary carc
12	158.5	15.4	446	22	ABB70063	Drosophila melanog
13	158	15.4	783	19	AAW37151	Mouse neural Mens+
14	158	15.4	787	19	AAW37152	Mouse neural Mens+
15	158	15.4	802	19	AAW37153	Mouse neural Mens+
16	158	15.4	802	22	AAU09139	Mammalian enabled
17	157	15.3	598	22	ABG14000	Novel human diagno
18	155	15.1	253	22	ABU52937	Human mammary carc
19	155	15.1	572	18	AAW31855	Mycobacterium tube
20	155	15.1	763	18	AAW31852	Mycobacterium tube
21	154.5	15.0	731	22	AAW74209	Protein encoded by
22	154.5	15.0	731	23	ABW33202	Herbicidally activ
23	153.5	14.9	177	22	ABU52946	Human mammary carc
24	151.5	14.7	441	22	ABB11413	Human extensin hom
25	151	14.7	693	23	ABP69529	Human polypeptide
26	150.5	14.6	533	16	AAR72483	Human H-2RIIBP. H
27	150.5	14.6	533	23	ABB79960	Retinoid C recepto
28	150.5	14.6	533	23	AAO19271	Human retinoid X r
29	150.5	14.6	533	23	AAO19282	Human retinoid X r
30	150.5	14.6	533	23	AAO19283	Human retinoid X r
31	150.5	14.6	533	23	AAO19284	Human retinoid X r
32	150.5	14.6	533	23	AAO19285	Human retinoid X r
33	150.5	14.6	533	23	AAO19286	Human retinoid X r
34	150.5	14.6	533	23	AAO19287	Human retinoid X r
35	150.5	14.6	533	23	AAO19288	Human retinoid X r
36	150.5	14.6	533	23	AAO19289	Human retinoid X r
37	150.5	14.6	533	23	AAO19291	Human retinoid X r
38	150.5	14.6	533	23	AAO19292	Human retinoid X r
39	150.5	14.6	533	23	AAO19293	Human retinoid X r
40	150.5	14.6	533	23	AAO19294	Human retinoid X r
41	150.5	14.6	533	23	AAO19295	Human retinoid X r
42	150.5	14.6	533	23	AAO19296	Human retinoid X r
43	150.5	14.6	533	23	AAO19297	Human retinoid X r
44	150.5	14.6	533	23	AAO19298	Human retinoid X r
45	150.5	14.6	533	23	AAO19299	Human retinoid X r

#### ALIGNMENTS

RESULT 1  
AAB71229  
ID AAB71229 standard; Protein; 1426 AA.

XX AAB71229;

XX 18-NOV-2002 (first entry)

XX Human legless homologue lgs/bcl9 protein.

XX Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
XX tissue proliferation; tumour; cytostatic; cellular disorder; colon;  
XX blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
XX medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX Homo sapiens.

XX US2002086986-A1.

XX 04-JUL-2002.

XX 27-JUL-2001; 2001US-0915543.

XX 28-JUL-2000; 2000US-221502P.

XX (BASL/) BASLER K.

XX (BRUN/) BRUNNER B.

XX (FROE/) FROESCH B.

XX (KRAM/) KRAMPS T.

XX (PETE/) PETER O.

XX Basler K, Brunner B, Froesch B, Kramps T, Peter O;

```

XX DR WPI; 2002-635689/68.
XX DR N-PSDB; AAF88467.
XX PT Novel polypeptide useful in therapeutic method for treating disorders
XX PT of cell fate such as cell differentiation or cell proliferation -
XX XX
XX PS Example II; Fig 8B; 41pp; English.
XX XX
XX CC This invention describes a novel polypeptide sharing one or more
XX CC homologous amino acid domains with the legless (lgs) protein, a
XX CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
XX CC involved in the formation and maintenance of spatial arrangements
XX CC and proliferation of tissues during development, and in the formation
XX CC and growth of many human tumours. The products of the invention have
XX CC cytotactic activity and can be used to treat cellular disorders, blood
XX CC disorders and cancers caused by over-stimulation of the Wnt pathway,
XX CC where the cancerous condition is colon, breast, head and neck, brain,
XX CC thyroid, medulloblastoma or skin cancer. The product could also be used
XX CC to promote tissue regeneration and repair. This sequence represents the
XX CC human legless (lgs) protein homologue lgs/bcl9 described in the
XX CC disclosure of the invention.
XX SQ Sequence 1426 AA;
    Query Match 100.0%; Score 1028; DB 23; Length 1426;
    Best Local Similarity 100.0%; Pred. No. 7.5e-65;
    Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TIVSHIQNISNNKTERSTAPLNTQISALRNDPKLPQQPPAPANQONSSQNTRLQPTP 60
DB 199 TIVSHIQNISNNKTERSTAPLNTQISALRNDPKLPQQPPAPANQONSSQNTRLQPTP 258
QY 61 PIPAPAPKPAAPPRPLDRESFGVENKLIPIVSGSPASSTPLPDGTGPNSTPNRAVTPVS 120
DB 259 PIPAPAPKPAAPPRPLDRESFGVENKLIPIVSGSPASSTPLPDGTGPNSTPNRAVTPVS 318
QY 121 QGSSSSADPKAPPPPPVSSGEPPTLGENPDGLSQGLEHRSQTLRIQRMFLPDEK 180
DB 319 QGSSSSADPKAPPPPPVSSGEPPTLGENPDGLSQGLEHRSQTLRIQRMFLPDEK 378
QY 181 EFTGAQSGGPOQNP 194
DB 379 EFTGAQSGGPOQNP 392
RESULT 2
ID ABB11808
XX ID ABB11808 standard; peptide; 1435 AA.
XX AC ABB11808;
XX DT 11-JAN-2002 (first entry)
XX DE Human BCL9 homologue, SEQ ID NO:2178.
XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX KW chronic inflammatory condition; proliferative retinopathy;
XX KW atherosclerosis; coronary heart disease; arterial ischaemia;
XX KW bone disorder; osteoporosis; vascular growth disorder;
XX KW tissue regeneration; wound healing; infection; immune disorder;
XX KW cell culture; drug screening; gene therapy; antiinflammatory;
XX KW antialasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
XX KW cytotactic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX KW antifungal; vulnetary; antiulcer.
XX OS Homo sapiens.
XX XX
XX PN WO200157188-A2.

```

```

XX PD 09-AUG-2001.
XX XX
XX PF 05-FEB-2001; 2001WO-US03800.
XX XX
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX XX
XX PI Tang YT, Liu C, Drmanac RT;
XX XX
XX DR WPI: 2001-457740/49.
XX DR N-PSDB; ABA09052.
XX XX
XX PT Human proteins and DNA encoding sequences useful for preventing,
XX PT treating or ameliorating a medical condition in a mammalian subject
XX PT e.g. arthritis and cancer -
XX XX
XX PS Claim 20; Page 256-257; 1963pp; English.
XX XX
XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
XX CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX CC invention also relates to vectors and recombinant host cells comprising a
XX CC nucleotide of the invention, methods of producing the novel polypeptides,
XX CC antibodies against the polypeptides, methods of detecting the nucleotides
XX CC or polypeptides in a sample, and methods of identifying compounds which
XX CC bind to polypeptides of the invention. Although novel, many of the
XX CC polypeptides of the invention have homology to known proteins, thereby
XX CC giving an insight into their probable biological activities. and hence
XX CC potential therapeutic applications. The polypeptides of the invention may
XX CC have various activities, including cytokine, cell proliferation or cell
XX CC differentiation activities; stem cell growth factor activity;
XX CC haematopoiesis regulatory activity; tissue growth activity;
XX CC immunomodulatory activity; activin- or inhibin-related activities;
XX CC chemotactic or chemokinetic activities; haemostatic; thrombotic or
XX CC thrombolytic activities; receptor or ligand activities; or may be
XX CC involved in oncogenesis, cancer cell proliferation or metastasis.
XX CC Depending on their biological activities, polypeptides and nucleotides of
XX CC the invention are useful for preventing, treating or ameliorating medical
XX CC conditions, e.g., by protein or gene therapy. Such conditions include
XX CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX CC proliferative retinopathy, atherosclerosis, coronary heart disease,
XX CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX CC vascular growth. Polypeptides involved with tissue regeneration and
XX CC repair (or nucleic acids encoding them) may be used to promote wound
XX CC healing (e.g., of burns, incisions and ulcers), while those with
XX CC immunomodulatory activities may be used in the treatment of viral,
XX CC bacterial and fungal infections in addition to immune disorders.
XX CC Polypeptides with growth factor activity may be used in cell cultures to
XX CC promote cell growth. For example, such polypeptides may be used to
XX CC manipulate stem cells in culture to give rise to neuroepithelial cells
XX CC that can be used to augment or replace cells damaged by illness,
XX CC autoimmune disease or accidental damage. The polypeptides and nucleotides
XX CC may also be used in the diagnosis of the above conditions, and in drug
XX CC screening techniques. The present sequence represents a novel human
XX CC polypeptide of the invention.
XX SQ Sequence 1435 AA;
    Query Match 100.0%; Score 1028; DB 22; Length 1435;
    Best Local Similarity 100.0%; Pred. No. 7.6e-65;
    Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TIVSHIQNISNNKTERSTAPLNTQISALRNDPKLPQQPPAPANQONSSQNTRLQPTP 60
DB 239 TIVSHIQNISNNKTERSTAPLNTQISALRNDPKLPQQPPAPANQONSSQNTRLQPTP 298
QY 61 PIPAPAPKPAAPPRPLDRESFGVENKLIPIVSGSPASSTPLPDGTGPNSTPNRAVTPVS 120
DB 299 PIPAPAPKPAAPPRPLDRESFGVENKLIPIVSGSPASSTPLPDGTGPNSTPNRAVTPVS 358

```

QY 121 QGNSSSADPKAPPPVSSGEPPTLGENPDGLSQLEHRSLSQTLRDIOHMLFPDEK 180  
 DB 359 QGNSSSADPKAPPPVSSGEPPTLGENPDGLSQLEHRSLSQTLRDIOHMLFPDEK 418  
 QY 181 EFTGAQSGGQQNP 194  
 DB 419 EFTGAQSGGQQNP 432

## RESULT 3

AAO05855  
 ID AAO05855 standard; Protein; 140 AA.

AC AAO05855;

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 19747.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI85786.

XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -

XX Claim 20; SEQ ID NO 19747; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 140 AA;

Query Match 26.9%; Score 277; DB 22; Length 140;

Best Local Similarity 91.4%; Pred. No. 1.7e-12;

Matches 53; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 TIVSFHIONISNNKTERSTAPLNTQISALRNDKPLFPQPPAPANQDQSSQNTLRQTP 58

DB 81 TWVSEHIONISSKTERSTAPLNTQISALRNDKPLFPQPPAPASQDQNSQNTLRQTP 138

## RESULT 4

AAU78461  
 ID AAU78461 standard; Protein; 320 AA.

XX AAU78461;

DT 02-JUL-2002 (first entry)

XX Mouse beta-catenin nuclear localised protein #2.

XX Mouse; beta-catenin nuclear localised protein; cancer;  
 KW gene therapy; EST; expressed sequence tag.

XX Mus musculus.

XX WO200224738-A1.

XX 28-MAR-2002.

PP 19-SEP-2001; 2001WO-JP08140.

PR 22-SEP-2000; 2000JP-0287876.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Akiyama T, Adachi S;

XX WPI; 2002-330014/36.

DR N-PSDB; ABK47632.

XX New beta-catenin nuclear localised protein for diagnosis and treatment  
 PT of diseases associated with nuclear localisation of beta-catenin e.g.  
 PT cancer -

XX Claim 2; Page 91-92; 113pp; Japanese.

XX The invention relates to a beta-catenin nuclear localised protein  
 CC and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC mouse beta-catenin nuclear localised protein #2.

XX Sequence 320 AA;

Query Match 22.6%; Score 232.5; DB 23; Length 320;

Best Local Similarity 35.1%; Pred. No. 5.7e-09;

Matches 71; Conservative 23; Mismatches 75; Indels 33; Gaps 9;

QY 1 TIVSFHIONISNNKTERSTAPLNTQISALRNDKPLFPQPPAPANQDQSSQNTLRQTP 60

DB 16 SILAYHQNVPRAKLQDA-----PKVPTPEPLPLN---TFSAGTFQSQPP 58

QY 61 PTPAPAPKPAAPRPLDRES- -GVENKLIP-SVGSPPASSTPLPDGTGPNSTPNRAVT 117

DB 59 PLPPLPPAPGSAAPALPEGPEDTSQDLAPNSVG- -AASGTGGTGGTHENTPTAATANN 116

QY 118 PVSQSSNSSADPKA- -PPPPVSSGEPPTLGENPDGLSQLEHRSLSQTLRDIOHML 175

DB 117 PLPFGDPSAPGSAALLGEATPTGNGQNLVGS- -EGLSKQLEHRSLSQTLRDIERLL 174

QY 176 F--PDEKFT- - - - -GAQSGGP 190

DB 175 LRSGETEFLKPGPGAGEGP 196

## RESULT 5

AAU78460

ID AAU78460 standard; Protein; 1494 AA.

XX AAU78460;

AC AAU78460;

DT 02-JUL-2002 (first entry)  
 XX Mouse beta-catenin nuclear localised protein.  
 XX Mouse; beta-catenin nuclear localised protein; cancer;  
 KW gene therapy; EST; expressed sequence tag.  
 XX Mus musculus.  
 OS  
 XX WO200224738-A1.  
 PN 28-MAR-2002.  
 PD  
 XX 19-SEP-2001; 2001WO-JP08140.  
 PF  
 XX 22-SEP-2000; 2000JP-0287876.  
 PR  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Akiyama T, Adachi S;  
 PI WPI; 2002-330014/36.  
 DR N-PSDB; ABK47631.  
 DR  
 XX New beta-catenin nuclear localised protein for diagnosis and treatment  
 PT of diseases associated with nuclear localisation of beta-catenin e.g.  
 PT cancer  
 PT  
 XX Claim 1; Page 81-88; 113pp; Japanese.  
 PS  
 XX The invention relates to a beta-catenin nuclear localised protein  
 CC and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC mouse beta-catenin nuclear localised protein.  
 CC  
 XX Sequence 1494 AA;  
 SQ  
 Query Match 22.6%; Score 232.5; DB 23; Length 1494;  
 Best Local Similarity 35.1%; Pred. No. 2.7e-08;  
 Matches 71; Conservative 23; Mismatches 75; Indels 33; Gaps 9;  
 QY 1 TIVSHQINISNKKTERSTAPLNTQISALRNDPKLPQOPAPANQONSSQNTLRQTP 60  
 DB 260 SILAYHQQNVRAKLDQA-----PKVPTFRPLPLN---TPRAGTFQSQPP 302  
 QY 61 PIPAPAPKPAAPRPLDRSP--GVENKLIP-SVGSASSTPLPPDGTGPNSTNNRAVT 117  
 DB 303 PLRPPPPAPGSAPPALPPEGPEDTSQDLAPNSVG--AASTGGGTGTHPTTAATANN 360  
 QY 118 PVSQSNSSSADPKA--PPPPVSSGEPPTLGENPDGLSQQLHRRSLQTLRDIQRL 175  
 DB 361 PLPPGSDPGSAPGSALLGATGTGNGRNLVGS--EGLSKQLHRRSLQTLRDLRL 418  
 QY 176 F--PDEKEPT----GAOSGGP 190  
 DB 419 LRSGTEPPLKGGPPGAGGGP 440  
 RESULT 6  
 AAM39141  
 ID AAM39141 standard; Protein; 707 AA.  
 XX  
 AC AAM39141;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 2286.  
 DE  
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

amytrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 leukaemia.  
 OS Homo sapiens.  
 XX WO200153312-A1.  
 PN 26-JUL-2001.  
 PD  
 XX 26-DEC-2000; 2000WO-US34263.  
 PF  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Dimañac RT;  
 XX WPI; 2001-442253/47.  
 DR N-PSDB; AAI58297.  
 DR  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PT  
 XX Example 4; SEQ ID NO 2286; 10078pp; English.  
 PS  
 XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AA42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 707 AA;  
 Query Match 16.3%; Score 167.5; DB 22; Length 707;  
 Best Local Similarity 40.0%; Pred. No. 0.00053;  
 Matches 54; Conservative 47; Mismatches 47; Indels 29; Gaps 8;  
 QY 30 RNDPKLPQOPAPANQON--SQNTLRQTPPLPAPAPKP-----AAPPLPDRSPGV 83  
 DB 53 QSGPKP-FIPPPPHQOQQQPPPPQPPPPHQPQPPHQPQHQOQQPPPPPPPOSSKPV 111  
 QY 84 EN--KLIPSVGS--PASSTPLP-----PDGTGPNSTPN-NRAVTPVSGSNSSSADP 130  
 DB 112 AQPGFAPGVGSAPASSAPATPTSGAPGSGGPTTPTTTPAVT-----SAPP 162  
 QY 131 KAPPPPVSSGEPPT 145  
 DB 163 GAPPTTSSGVPTT 177  
 RESULT 7  
 AAM52322

ID AAMS2322 standard; Protein: 574 AA.  
 AC AAM52322;  
 DT 18-JAN-2002 (first entry)  
 DE WASP homolog protein.  
 KW Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;  
 KW metastatic cancer; parasitic infection; cytotoxic; WASP.  
 OS Schizosaccharomyces pombe.  
 PN WO200171356-A2.  
 XX 27-SEP-2001.  
 XX 21-MAR-2001; 2001MO-FR00843.  
 XX 22-MAR-2000; 2000FR-0003637.  
 XX (CNRS ) CENT NAT RECH SCI.  
 XX (CURI-) INST CURIE.  
 XX Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V;  
 XX Sykes C;  
 XX WPI; 2001-639148/73.  
 XX Identifying modulators of actin polymerization, potentially useful for  
 PT treating tumor metastasis and parasitic infection, using proteins that  
 PT contain Ena/VASP binding sites -  
 XX Claim 13; Pages 107-109; 109pp; French.  
 XX The present invention relates to a method for identifying modulators of  
 CC actin polymerisation. The method involves using proteins that contain at  
 CC least one binding motif for proteins of the Ena/VASP  
 CC (vasodilator-stimulated phosphoprotein) family in the preparation of  
 CC reagents for identification/screening of molecules that modulate  
 CC formation of the actin cytoskeleton. The proteins used in the method  
 CC (i.e. the proteins with binding motifs) for Ena/VASP proteins) do not  
 CC bind to the Arp2/3 protein complex. The modulators identified by the  
 CC method are potentially useful for treating disorders of actin  
 CC polymerisation, e.g. metastatic cancer or parasitic infection, and as  
 CC cytotoxic agents. The present sequence one such protein with binding  
 CC motifs) for Ena/VASP proteins, which was used in the method of the  
 CC present invention.  
 XX SQ Sequence 574 AA;  
 Query Match 15.8%; Score 162.5; DB 22; Length 574;  
 Best Local Similarity 30.9%; Pred. No. 0.00097;  
 Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
 QY 7 IONISNNKTERSTA-----PLNTQISA---LRNDPKLPQPPAPANODQ-----NSSQNT 54  
 Db 275 IAPVSMFPAINSTKPLPFPSSRSVSAALAAANKKPPPPPPSRNRKPKPIGNGSSNS 334  
 QY 55 RLQPTTP-----IPAPAKPAAPRPLDRESGVENKLIPIVSGSPA-SSTPLPDG 104  
 Db 335 SLPPPPPPPSRNAAGSIPLPQGRSAPPPPPPPPSAPSTGRQPPFLSSRAVSNPPAPP 394  
 QY 105 TGNSTPNNRVAVTVSQQSSSSADPKAPPPPPVSSGPPPTL 146  
 Db 395 IPGRSAP---ALPPLGNASRTST--PPVTPPSLPPSAPP 431  
 RESULT 8  
 AAG67370  
 ID AAG67370 standard; Protein: 574 AA.  
 XX AAG67370;  
 AC AAG67370;

XX 13-NOV-2001 (first entry)  
 XX Amino acid sequence of a yeast WASP protein homologue.  
 XX Wiskott-Aldrich syndrome protein; WASP; actin cytoskeleton;  
 KW cell motility; actin polymerisation; cancer; parasite infection;  
 KW embryonic development; immune response; wound repair.  
 OS Schizosaccharomyces pombe.  
 PN WO200144292-A2.  
 XX 21-JUN-2001.  
 XX 15-DEC-2000; 2000WO-FR03569.  
 XX 16-DEC-1999; 99FR-0015900.  
 XX (CNRS ) CENT NAT RECH SCI.  
 XX (CURI-) INST CURIE.  
 XX Noireaux V, Prost J, Sykes C, Friederich E, Golsteyn RM;  
 XX Louvard D;  
 XX WPI; 2001-536241/59.  
 XX N-PSDB; AAH77922.  
 XX New fragments of WASP family proteins, useful for detecting and  
 PT identifying modulators of actin cytoskeleton formation, potential  
 PT anticancer and antiparasitic agents -  
 XX Claim 14; Fig 8; 162pp; French.  
 XX The present sequence represents a WASP (Wiskott-Aldrich syndrome  
 CC protein) homologue. Peptide fragments of WASP-family proteins of  
 CC eukaryotic cells are used to prepare reagents for detecting compounds  
 CC that inhibit or stimulate formation of the actin cytoskeleton, and  
 CC thus inhibit or stimulate cell motility. The peptides are used to  
 CC detect and identify compounds which are potentially useful for treating  
 CC diseases associated with dysfunction of actin polymerisation,  
 CC particularly metastatic cancer and parasite infection; as cytotoxic  
 CC agents for inhibiting/stimulating formation of the actin cytoskeleton  
 CC and for detecting side-effects, on actin polymerisation, of  
 CC pharmaceuticals. By modulating actin polymerisation, these compounds  
 CC affect cell motility, embryonic development, the immune response and  
 CC wound repair.  
 XX SQ Sequence 574 AA;  
 Query Match 15.8%; Score 162.5; DB 22; Length 574;  
 Best Local Similarity 30.9%; Pred. No. 0.00097;  
 Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
 QY 7 IONISNNKTERSTA-----PLNTQISA---LRNDPKLPQPPAPANODQ-----NSSQNT 54  
 Db 275 IAPVSMFPAINSTKPLPFPSSRSVSAALAAANKKPPPPPPSRNRKPKPIGNGSSNS 334  
 QY 55 RLQPTTP-----IPAPAKPAAPRPLDRESGVENKLIPIVSGSPA-SSTPLPDG 104  
 Db 335 SLPPPPPPPSRNAAGSIPLPQGRSAPPPPPPPPSAPSTGRQPPFLSSRAVSNPPAPP 394  
 QY 105 TGNSTPNNRVAVTVSQQSSSSADPKAPPPPPVSSGPPPTL 146  
 Db 395 IPGRSAP---ALPPLGNASRTST--PPVTPPSLPPSAPP 431  
 RESULT 9  
 ABG27250  
 ID ABG27250 standard; Protein: 406 AA.  
 XX ABG27250;  
 AC ABG27250;

DT 18-FEB-2002 (first entry)  
DE Novel human diagnostic protein #27241.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
OS WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-539362/73.  
DR N-PSDB; AAS91437.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 57609; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 406 AA;  
Query Match 15.5%; Score 159; DB 22; Length 406;  
Best Local Similarity 32.0%; Pred. No. 0.0012;  
Matches 41; Conservative 15; Mismatches 62; Indels 10; Gaps 4;  
QY 33 PKLFPQPPAPFA-----NQCNSSQNTLQPTTTPAPAPKFA-APPRLDRESFGVE 84  
DB 20 PSLPSP 79  
QY 85 N-KLIPVGSFASSTPLPDDGTGNSTP-NNRVTPVSGSNSSADPKAPPPPVSSGE 142  
DB 80 SP 139  
QY 143 PPTLGENP 150  
DB 140 PPSPPSP 147

RESULT 10  
ABUS2945  
ID ABUS2945 standard; Protein; 256 AA.  
XX AC ABUS2945;  
XX 14-APR-2003 (first entry)  
XX Human mammary carcinoma-derived DKFZphmcfl\_1c23 homologue #15.  
XX Human; gene therapy; vaccine; disease treatment; detection.  
XX Homo sapiens.  
XX WO200112659-A2.  
XX 22-FEB-2001.  
XX 18-AUG-2000; 2000WO-IB01496.  
XX 18-AUG-1999; 99US-0149499.  
PR 28-SEP-1999; 99US-0156503.  
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX Wiemann S;  
XX WPI; 2001-327840/34.  
XX Nucleic acids having the sequences of clones isolated from libraries of  
XX different human tissues, useful in recombinant DNA methodologies -  
XX Example III; Page 551; 1095pp; English.  
XX This invention describes novel polynucleotides and polypeptides isolated  
XX from human DNA libraries which can be used for gene therapy or in  
XX vaccines. The polynucleotides of the invention and antibodies encoded by  
XX them may be used in the prevention, diagnosis and treatment of diseases  
XX associated with inappropriate polypeptide expression. The products of the  
XX invention may also be used to identify modulators of expression and  
XX activity and to down regulate expression and activity. The antibodies of  
XX the invention may also be used as diagnostic agents for detecting the  
XX presence of polypeptides in samples. This sequence represents a homologue  
XX of a polypeptide described in the disclosure of the invention.  
XX SQ Sequence 256 AA;  
Query Match 15.4%; Score 158.5; DB 22; Length 256;  
Best Local Similarity 32.0%; Pred. No. 0.00082;  
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;  
QY 14 KTERSTAPLNTQISALRNDPKLP--QPPAPANQDNSSQNTLQPP---TPPIAPAPK 68  
DB 46 KSSPPAPVNLPPPEVKSPPPTPVSSPPAPKSPPPPMSSPPPEVKSPPPPAPVSS 105  
QY 69 P-----AAPRLDRESFQVENKLIPVSGSFASSTPLP---PDGTGNSTPNRAVTPV 119  
DB 106 PPEPVKSPPPPAPVSGSPPPPVKS---PPPPAPVSSPPPPVKSPPPPAPVSSPPPVKSPP 162  
QY 120 SQGSNSSADP-KAPPPP-PVSSGEPPTLGENP 150  
DB 163 PPAVSSPPPPVKSPPPPAPVSSPPPPPIKSPPP 195  
RESULT 11  
ABUS2939  
ID ABUS2939 standard; Protein; 270 AA.  
XX AC ABUS2939;  
XX 14-APR-2003 (first entry)  
XX Human mammary carcinoma-derived DKFZphmcfl\_1c23 homologue #9.

XX KW Human; gene therapy; vaccine; disease treatment; detection.  
 XX OS Homo sapiens.  
 XX PN WO200112659-A2.  
 XX PD 22-FEB-2001.  
 XX PF 18-AUG-2000; 2000WO-IB01496.  
 XX PR 18-AUG-1999; 99US-0149499.  
 XX PR 28-SEP-1999; 99US-0156503.  
 XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 XX PI Wiemann S;  
 XX DR WPI; 2001-327840/34.  
 XX CC Nucleic acids having the sequences of clones isolated from libraries of  
 PT different human tissues, useful in recombinant DNA methodologies -  
 XX PS Example III; Page 550; 1095pp; English.  
 XX CC This invention describes novel polynucleotides and polypeptides isolated  
 CC from human cDNA libraries which can be used for gene therapy or in  
 CC vaccines. The polynucleotides of the invention and antibodies encoded by  
 CC them may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression. The products of the  
 CC invention may also be used to identify modulators of expression and  
 CC activity and to down regulate expression and activity. The antibodies of  
 CC the invention may also be used as diagnostic agents for detecting the  
 CC presence of polypeptides in samples. This sequence represents a homologue  
 CC of a polypeptide described in the disclosure of the invention.  
 XX SQ Sequence 270 AA;  
 Query Match 15.4%; Score 158.5; DB 22; Length 270;  
 Best Local Similarity 32.0%; Pred. No. 0.00087;  
 Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;  
 QY 14 KTERSTAPLNTQISALRNDKPLP--QQPPAPANQDNSSQNTLQP---TPPIAPAPK 68  
 Db 79 KSPPAPVNLPPPEVKSSPPTPVSSPPFPAPKSSPPFPAPMSPPPPEVKSPPPAPVSS 138  
 QY 69 P-----AAPRLDRESGVENKLIPSVGSPASSTPLP---PDGTPNSTPNNRAVTPV 119  
 Db 139 PPPPVKSPPPAPVSSPPPEVKSS---PPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSP 195  
 QY 120 SQGSNSSSDRP-KAPPPP-PVSSGEPPTLGENP 150  
 Db 196 PPAPVSSPPPPVKSPPPPPAPVSSPPPPPKSPPP 228  
 RESULT 12  
 ABB70063  
 ID ABB70063 standard; Protein; 446 AA.  
 XX AC ABB70063;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 36981.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX OS pharmaceutical.  
 XX PN Drosophila melanogaster.  
 XX PD WO200171042-A2.  
 PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 XX 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL14166.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX PS Disclosure; SEQ ID NO 36981; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABBS7737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 446 AA;  
 Query Match 15.4%; Score 158.5; DB 22; Length 446;  
 Best Local Similarity 29.8%; Pred. NO. 0.0014;  
 Matches 42; Conservative 11; Mismatches 64; Indels 25; Gaps 3;  
 QY 33 KPFLPQQPPAPANQDNSSQNTLQPPTPIAPA--PKPAAPPRLDRESGVENKLIPS 90  
 Db 97 PQRTPPPPPPP-----QTPPAPRPSYGPQTQTPPPPPPTPSAPAPPPPS 144  
 QY 91 VG-----SPASSTPLPPDGTGPNSTPNNRAVTPVSGSNSSSADPKAPPPPPVS 139  
 Db 145 YGPPQTPPP 204  
 QY 140 SGEPTLGENPDGLSQEQLEHR 161  
 Db 205 PPSPPSPQPGPYLPPDQPKPR 226  
 RESULT 13  
 AAW37151  
 ID AAW37151 standard; Protein; 783 AA.  
 XX AC AAW37151;  
 XX DT 06-JUL-1998 (first entry)  
 XX DE Mouse neural Mena+ protein.  
 XX KW Neural Mena+ protein; mammalian Ena; Enabled protein; Evi protein;  
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;  
 KW cell growth; cell motility; mouse.  
 XX OS Mus musculus.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 378  
 FT /note= "encoded by GGN"  
 XX PN WO9801755-A1.  
 XX PD 15-JAN-1998.

XX PF 03-JUL-1997; 97WO-US11669.  
XX PR 05-JUL-1996; 96US-0675815.  
XX PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
XX PI Gertler FB, Niebuhr K, Soriano P, Wehland J;  
XX DR WPI; 1998-101197/09.  
XX DR N-PSDB; AAV02998.  
XX PT Detection of modulators of Mena and Ena-VASP-like genes and proteins  
XX PT - used in control of cytoskeletal dynamic events in normal and  
XX PT abnormal cell morphology, adhesion, motility, growth and  
XX PT differentiation  
XX PS Example 4; Page 58-60; 77pp; English.  
XX CC This protein comprises novel murine neural Mena+. Its amino acid  
XX CC sequence was deduced from a cDNA clone (see AAV02998) obtained from  
XX CC a mouse brain cDNA library. Neural Mena+ contains an exon that  
XX CC introduces 244 amino acids between amino acids 238 and 239 of  
XX CC mammalian Ena (Mena, see AAW37148). Two other isoforms, neural  
XX CC Mena++ (see AAW37152) and neural Mena+++ (see AAW37153), are also  
XX CC disclosed. Unlike Mena, neural Mena isoforms exhibit neural  
XX CC tissue-specific distribution. Based on the disclosed Mena and Evi  
XX CC genes (see also AAV02996-97) and proteins (see also AAW37148-49), a  
XX CC variety of methods and compositions are provided for screening,  
XX CC isolating and characterising endogenous and exogenous factors,  
XX CC drugs and therapeutic agents useful to evaluate and/or control  
XX CC cytoskeletal dynamic events involved in normal and abnormal cell  
XX CC morphology, adhesion, motility, growth and/or differentiation. A  
XX CC method of detecting a modulator of Mena activity/expression is  
XX CC claimed.  
XX SQ Sequence 783 AA;  
Query Match 15.4%; Score 158; DB 19; Length 783;  
Best Local Similarity 26.3%; Pred. No. 0.0028;  
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;  
QY 13 NKTERSTAPLNT-----QISALRNDP----- 33  
DB 329 NKNRSPSPVNTSSQPPAAKSCAWPTSNFSLPPSPFPMISSPPGKATGPRVLPVCVS 388  
QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTTPIAPAPKPAAPRPL----- 76  
DB 389 SPVQMPSPPTAPNGSLDSVTYPVSPPTTSGPAAAPPPPPPPPPPLPPLPLPLAS 448  
QY 77 -----DRESRGVENKLI-----PSVGSPPASS-TPLPPD-GTGNSTPNNAV- 116  
DB 449 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSAGAPASAEPLNPELGDSSASEPGLQAAAS 508  
QY 117 ----TPVSGSNSSSADPKAPPPPPVSSGE-----PPTLGENP 150  
DB 509 QPAESPTFG--LVLGPPAPPPPLPSGAYASALFPFPPGPP 550  
RESULT 14  
AAW37152  
ID AAW37152 standard; Protein; 787 AA.  
XX AC AAW37152;  
XX AC AAW37153;  
XX DT 06-JUL-1998 (first entry)  
XX DE Mouse neural Mena++ protein.  
XX KW Neural Mena++ protein; mammalian Ena; Enabled protein; Evi protein;  
XX KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;  
XX KW cell growth; cell motility; mouse.

XX OS Mus musculus.  
XX FN WO9801755-A1.  
XX XX 15-JAN-1998.  
XX PD 03-JUL-1997; 97WO-US11669.  
XX PF 05-JUL-1996; 96US-0675815.  
XX PR (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
XX PA Gertler FB, Niebuhr K, Soriano P, Wehland J;  
XX DR WPI; 1998-101197/09.  
XX PT Detection of modulators of Mena and Ena-VASP-like genes and proteins  
XX PT - used in control of cytoskeletal dynamic events in normal and  
XX PT abnormal cell morphology, adhesion, motility, growth and  
XX PT differentiation  
XX PS Example 4; Page 60-63; 77pp; English.  
XX CC This protein comprises novel murine neural Mena+. Its amino acid  
XX CC sequence was deduced from a cDNA clone obtained from a mouse brain  
XX CC cDNA library. Two other isoforms, neural Mena+ (see AAW37151) and  
XX CC neural Mena+++ (see AAW37153), are also disclosed. Unlike mammalian  
XX CC Ena (Mena, see AAW37148), neural Mena isoforms exhibit neural  
XX CC tissue-specific distribution. Based on the disclosed Mena and Evi  
XX CC genes (see also AAV02996-98) and proteins (see also AAW37148-49), a  
XX CC variety of methods and compositions are provided for screening,  
XX CC isolating and characterising endogenous and exogenous factors,  
XX CC drugs and therapeutic agents useful to evaluate and/or control  
XX CC cytoskeletal dynamic events involved in normal and abnormal cell  
XX CC morphology, adhesion, motility, growth and/or differentiation. A  
XX CC method of detecting a modulator of Mena activity/expression is  
XX CC claimed.  
XX SQ Sequence 787 AA;  
Query Match 15.4%; Score 158; DB 19; Length 787;  
Best Local Similarity 26.3%; Pred. No. 0.0028;  
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;  
QY 13 NKTERSTAPLNT-----QISALRNDP----- 33  
DB 333 NKNRSPSPVNTSSQPPAAKSCAWPTSNFSLPPSPFPMISSPPGKATGPRVLPVCVS 392  
QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTTPIAPAPKPAAPRPL----- 76  
DB 393 SPVQMPSPPTAPNGSLDSVTYPVSPPTTSGPAAAPPPPPPPPPPLPPLPLPLAS 452  
QY 77 -----DRESRGVENKLI-----PSVGSPPASS-TPLPPD-GTGNSTPNNAV- 116  
DB 453 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSAGAPASAEPLNPELGDSSASEPGLQAAAS 512  
QY 117 ----TPVSGSNSSSADPKAPPPPPVSSGE-----PPTLGENP 150  
DB 513 QPAESPTFG--LVLGPPAPPPPLPSGAYASALFPFPPGPP 554  
RESULT 15  
AAW37153  
ID AAW37153 standard; Protein; 802 AA.  
XX AC AAW37153;  
XX AC AAW37153;  
XX DT 06-JUL-1998 (first entry)  
XX DE Mouse neural Mena+++ protein.



KW Neural Mena+++ protein; mammalian Ena; Enabled protein; Evi protein;  
KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;  
KW cell growth; cell motility; mouse.

OS Mus musculus.

PN WO9801755-A1.

PO 15-JAN-1998.

XX 03-JUL-1997; 97WO-0511669.

XX 05-JUL-1996; 96US-0675815.

XX (GBPB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX Gertler FB, Niebuhr K, Soriano P, Wehland J;

XX WPI; 1998-101197/09.

XX Detection of modulators of Mena and Ena-VASP-like genes and proteins  
PT - used in control of cytoskeletal dynamic events in normal and  
PT abnormal cell morphology, adhesion, motility, growth and  
PT differentiation

XX Example 4; Page 63-65; 77pp; English.

XX This protein comprises novel murine neural Mena+-. Its amino acid  
CC sequence was deduced from a cDNA clone obtained from a mouse brain  
CC cDNA library. Two other isoforms, neural Mena+ (see AAW37151) and  
CC neural Mena++ (see AAW37152), are also disclosed. Unlike mammalian  
CC Ena (Mena, see AAW37148), neural Mena isoforms exhibit neural  
CC tissue-specific distribution. Based on the disclosed Mena and Evi  
CC genes (see also AAV02996-98) and proteins (see also AAW37148-49), a  
CC variety of methods and compositions are provided for screening,  
CC isolating and characterising endogenous and exogenous factors,  
CC drugs and therapeutic agents useful to evaluate and/or control  
CC cytoskeletal dynamic events involved in normal and abnormal cell  
CC morphology, adhesion, motility, growth and/or differentiation. A  
CC method of detecting a modulator of Mena activity/expression is  
CC claimed.

XX Sequence 802 AA;

Query Match 15.4%; Score 158; DB 19; Length 802;  
Best Local Similarity 26.3%; Pred. No. 0.0028;  
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;  
QY 13 NKTERSTAPLNT-----QISLRNDP-----33  
DB 348 NNNRSPSPVNTFSSQPPAAKSCAMPTSNFSLPSPIMISSPGKATGRPVLPVCVS 407  
QY 34 KFLPQQPPAPANQDN-----SSNTRLOPTTPIPAPAPKPAAPRPL-----76  
DB 408 SEVPQMPSPSTAPNGSLDSVTYV8PPPTSGPAAPPPPPPPPPPPPPPLPLAS 467  
QY 77 -----DRESPGVENKLI-----PSVGSFSS--TLPED-GTGNSTNNRAV- 116  
DB 468 LSHCGSQAPPPGTPLASTPSSKSPVLPSPSAGAPASAEPLNPLGSSASSEPLQAA 527  
QY 117 -----TPVQSGNSSSADKAPPPPPVSSGE-----PPTLGENP 150  
DB 528 QPAESPTQ3--LVLPAPAPPPPPPLPSGPAVASALPPPPPPP 569

Search completed: November 13, 2003, 09:25:18  
Job time : 80.2607 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:19 ; Search time 28.6848 Seconds  
(without alignments)  
286.155 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392  
Perfect score: 1028  
Sequence: 1 TIVSFHQMISNKKTERSTA.....LPPDEKFTGAQSGGPGQNP 194

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	161.5	15.7	816	4	US-09-266-225D-12
2	150.5	14.6	533	1	US-07-952-800-2
3	149.5	14.5	525	3	US-08-764-870-7
4	149.5	14.5	525	3	US-08-980-115-7
5	143.5	14.0	446	1	US-07-952-800-4
6	140.5	13.7	8991	4	US-08-714-741-32
7	138.5	13.5	297	2	US-08-580-545B-6
8	138.5	13.5	297	3	US-09-262-653A-6
9	137	13.3	456	2	US-08-819-013-1
10	136.5	13.3	142	4	US-09-252-991A-24873
11	136.5	13.3	1248	2	US-09-080-897-2
12	136.5	13.3	1248	3	US-09-321-735-2
13	135	13.1	332	3	US-08-818-112-53
14	135	13.1	332	4	US-08-818-111-53
15	135	13.1	332	4	US-09-056-556-53
16	135	13.1	332	4	US-09-072-596-53
17	135	13.1	334	6	5202236-3
18	133.5	13.0	941	4	US-07-757-022B-14
19	133.5	13.0	1022	4	US-07-757-022B-84
20	133.5	13.0	1038	4	US-07-757-022B-74
21	133.5	13.0	1049	4	US-07-757-022B-58
22	133.5	13.0	1140	4	US-07-757-022B-104
23	133.5	13.0	1270	4	US-07-757-022B-44
24	133.5	13.0	1311	4	US-07-757-022B-42
25	133.5	13.0	1313	4	US-07-757-022B-142
26	133.5	13.0	1314	4	US-07-757-022B-50
27	133.5	13.0	1320	4	US-07-757-022B-46

28 133.5 13.0 1320 4 US-07-757-022B-60 Sequence 60, Appl  
29 133.5 13.0 1354 4 US-07-757-022B-48 Sequence 48, Appl  
30 133.5 13.0 1361 4 US-07-757-022B-40 Sequence 40, Appl  
31 133.5 13.0 1363 4 US-07-757-022B-52 Sequence 52, Appl  
32 133.5 13.0 1404 4 US-07-757-022B-2 Sequence 2, Appl  
33 133.5 13.0 1404 4 US-07-757-022B-62 Sequence 62, Appl  
34 133 12.9 1315 3 US-08-899-595-3 Sequence 3, Appl  
35 132 12.8 214 1 US-08-217-327-4 Sequence 4, Appl  
36 132 12.8 331 6 5202236-37 Patent No. 5202236  
37 132 12.8 1231 4 US-08-714-741-41 Sequence 41, Appl  
38 131 12.7 2442 4 US-09-514-247A-10 Sequence 10, Appl  
39 130.5 12.7 325 1 US-08-382-184-2 Sequence 2, Appl  
40 130.5 12.7 325 3 US-08-641-356-2 Sequence 2, Appl  
41 130.5 12.7 325 3 US-09-132-528-2 Sequence 2, Appl  
42 130.5 12.7 325 3 US-09-132-528-3 Sequence 2, Appl  
43 130.5 12.7 325 3 US-08-875-494-2 Sequence 2, Appl  
44 130.5 12.7 325 4 US-09-599-366-2 Sequence 2, Appl  
45 130.5 12.7 325 4 US-09-599-366-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-09-266-225D-12  
; Sequence 12, Application US/09266225D  
; Patent No. 6573364  
; GENERAL INFORMATION:  
; APPLICANT: Nandabalan, Krishan  
; APPLICANT: Kingmore, Stephen  
; APPLICANT: Tchernyev, Vellizar  
; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak  
; TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Protein-  
; TITLE OF INVENTION: Interacting Proteins  
; FILE REFERENCE: 15966-523  
; CURRENT APPLICATION NUMBER: US/09/266, 225D  
; CURRENT FILING DATE: 1999-03-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 816  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-266-225D-12

Query Match 15.7%; Score 161.5; DB 4; Length 816;  
Best Local Similarity 32.1%; Pred. No. 1.6e-05;  
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;  
QY 17 RSTAPLNTQISALRNDFKPLP-----QQPPAPANDQDQSSQNTLRLOTPPIP 63  
DB 577 RPAPALTSVPAPAPATPTTTPVQPTSPPPGVAQPTGPPQPSAGSTSGVPQPACPPP 636  
QY 64 APAPKPAAPRPRLDRSPG-----VENKLIPSVGSPASSTP-----LPPDGT 105  
DB 637 GPAPHTGPPGPIVPAPAPQIATSTSLAAQSLVPPPLGSGSTPGVLPYFPFPLPPDA 696  
QY 106 G--PNS-----TNNRAVTPVSGSSNSADPKAPPVPPVSGEPPTLG 147  
DB 697 GGAPQSSMSSEPDNLVT--QLSKSQVEDFL----PPVFGTGTCKSG 738

RESULT 2  
US-07-952-800-2  
; Sequence 2, Application US/07952800  
; Patent No. 5403925  
; GENERAL INFORMATION:  
; APPLICANT: OZATO, KENZO  
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE  
; TITLE OF INVENTION: RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND

STREET: 1 MARKET PLAZA, STEUART TOWER, SUITE 2000  
CITY: SAN FRANCISCO  
STATE: CA  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/952,800  
FILING DATE: 19920928  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Kenneth A  
REGISTRATION NUMBER: 31,677  
REFERENCE/DOCKET NUMBER: 15280-21-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-543-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 533 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-952-800-2

Query Match 14.6%; Score 150.5; DB 1; Length 533;  
Best Local Similarity 31.0%; Pred. No. 7.7e-05;  
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;  
QY 31 NDPKPLPQ--QPPAPANQDQNSQNTLRQTPPIPA-----PKPAAPPPLDRESP 81  
DB 86 SSPNPLQGVPPSPPG-----PPLPSTAPSLGSGAGAPPMPPPPLGSPFP 134  
QY 82 GVENKLIPSVGSPASSPLPPDGTGPNSTNNRAVTPVSGSNSSADPKAP-----133  
DB 135 VISS-----SMGSPGLPPAPPFGSPVSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 190  
QY 134 -PPPVSSG 141  
DB 191 CPPPPGGFG 199

RESULT 3  
US-08-764-870-7  
Sequence 7, Application US/08764870  
Patent No. 6235946  
GENERAL INFORMATION:  
APPLICANT: Scanlan, Thomas S  
APPLICANT: Baxter, John D  
APPLICANT: Fletcher, Robert J  
APPLICANT: Wagner, Richard L  
APPLICANT: Kushner, Peter J  
APPLICANT: Apriletti, James W  
APPLICANT: West, Brian  
TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand  
TITLE OF INVENTION: Binding Domains  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESS: Cooley Godward  
STREET: Five Palo Alto Square, 3000 El Camino Real  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,870  
FILING DATE: 13-DEC-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/008,540  
FILING DATE: 13-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/008,543  
FILING DATE: 13-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/008,606  
FILING DATE: 14-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakamura, Jackie N  
REGISTRATION NUMBER: 35,966  
REFERENCE/DOCKET NUMBER: UCAL-246/01US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 843-5000  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-764-870-7

Query Match 14.5%; Score 149.5; DB 3; Length 525;  
Best Local Similarity 31.0%; Pred. No. 9.1e-05;  
Matches 40; Conservative 10; Mismatches 46; Indels 33; Gaps 5;  
QY 31 NDPKPLPQ--QPPAPANQDQNSQNTLRQTPPIPA-----PKPAAPPPLDRESP 81  
DB 78 SSPNPLQGVPPSPPG-----PPLPSTAPSLGSGAGAPPMPPPPLGSPFP 126  
QY 82 GVENKLIPSVGSPASSPLPPDGTGPNSTNNRAVTPVSGSNSSADPKAP-----133  
DB 127 VISS-----SMGSPGLPPAPPFGSPVSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 182  
QY 134 -PPPVSSG 141  
DB 183 CPPPPGGFG 191

RESULT 4  
US-08-980-115-7  
Sequence 7, Application US/08980115  
Patent No. 6266622  
GENERAL INFORMATION:  
APPLICANT: Scanlan, Thomas S  
APPLICANT: Baxter, John D  
APPLICANT: Fletcher, Robert J  
APPLICANT: Wagner, Richard L  
APPLICANT: Kushner, Peter J  
APPLICANT: Apriletti, James W  
APPLICANT: West, Brian L  
APPLICANT: Shiao, Andrew K  
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS  
TITLE OF INVENTION: UCAL-246/02US  
FILE REFERENCE: UCAL-246/02US  
CURRENT APPLICATION NUMBER: US/08/980,115  
CURRENT FILING DATE: 1997-11-26  
EARLIER APPLICATION NUMBER: 08/764,870  
EARLIER FILING DATE: 1996-12-13  
EARLIER APPLICATION NUMBER: 60/008,606  
EARLIER FILING DATE: 1995-12-14  
EARLIER APPLICATION NUMBER: 60/008,543  
EARLIER FILING DATE: 1995-12-13  
EARLIER APPLICATION NUMBER: 60/008,540  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7

```
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (292)..(523)
; OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-7

Query Match 14.5%; Score 149.5; DB 3; Length 525;
Best Local Similarity 31.0%; Pred. No. 9.1e-05;
Matches 40; Conservative 10; Mismatches 46; Indels 33; Gaps 5;

Qy 31 NDKPLPQ--QPAPANQDNSSQNTLQPTTPIPA-----PKPAAPRPLDRESP 81
Db 78 SSNPLPQGVPPSPG-----PPLPSTAPTLGGSGAPPPMPPLGSPFP 126
Qy 82 GVENKLIPIVSGSPASSTPLPDGTGPNSTNNRAVTVSQQNSSSADPKAP-----133
Db 127 VISS---SMGSPGLPPAPPFGFSPVSSPQINSTVSLPGGSGPPEDVKPVLGVRGLH 182
Qy 134 -PPPPVSSG 141
Db 183 CFPPPGPG 191

RESULT 5
US-07-952-800-4
; Sequence 4, Application US/07952800
; Patent No. 5403925
; GENERAL INFORMATION:
; APPLICANT: OZATO, KEIKO
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE
; TITLE OF INVENTION: RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND
; STREET: 1 MARKET PLAZA, STEUART TOWER, SUITE 2000
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,800
; FILING DATE: 19920928
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-21-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-952-800-4

Query Match 14.0%; Score 143.5; DB 1; Length 446;
Best Local Similarity 30.3%; Pred. No. 0.00023;
Matches 37; Conservative 12; Mismatches 44; Indels 29; Gaps 4;

Qy 31 NDKPLPQ--QPAPANQDNSSQNTLQPTTPIPAKPAAPRPLDRESPGVENKLI 88
Db 9 SSNPLSQGIRPSP-----PGPLTFSAAPPMPPLGSPFPVSS---51

; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (292)..(523)
; OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-7

Query Match 13.7%; Score 140.5; DB 4; Length 8991;
Best Local Similarity 27.4%; Pred. No. 0.014;
Matches 43; Conservative 9; Mismatches 62; Indels 43; Gaps 5;

Qy 16 ERSTAPLNTQISALRND-----PKPLPQPPAPANQDNSSQNTLQPTP--PIP 63
Db 5961 EKTAAEKKAELKAEADLKAVDEPTTAPAPAPAPA-----PTDPAPAP 6007
Qy 64 APAPKPAAPRPLDRESPGVENKLIPIVSGSPASSTPLPDGTGPNSTNNRAVTVSQQS 123
Db 6008 APAPKPAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 6055
Qy 124 NSSSADPKAPPPPPVSSGPPPTLGENPD-----GLSQE 156
Db 124 NSSSADPKAPPPPPVSSGPPPTLGENPD-----GLSQE 156
```



```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/788,322
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-64383-1/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-819-013-1

Query Match 13.3%; Score 137; DB 2; Length 456;
Best Local Similarity 25.9%; Pred. No. 0.00083;
Matches 51; Conservative 26; Mismatches 58; Indels 62; Gaps 10;

QY 10 ISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANODQSSNTRLQPTPPIPAPAPKP 69
DB 120 IDNRSSQRHSPFFS-----KTLPSKPSWP-----SEKARLTSLPALTALQXP 162
QY 70 AAPRP---LDRESPGV-----ENKLIPSVGS---PASSTPLPPDGTGPNSTPNNAVT 117
DB 163 QVPPKPKGLLDEADYVVFVDNDENIHTTESSPPPEKAPWNRSTKNS-----T 216
QY 118 PVS-----QGSNSSADPKAPPP-----PPVSSGGPPTL-----GPNGLGLSEQ-- 157
DB 217 PASPGTASGRNSGAWETKSPPAAPSPLPRAGKPTTLPKTPVASQSNASSVCSEKPI 276
QY 158 -----LEHRSLOT 167
DB 277 PAERHGRSSHRQEAQVS 293

RESULT 10
US-09-252-991A-24873
; Sequence 24873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24873
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24873

Query Match 13.3%; Score 136.5; DB 4; Length 142;
Best Local Similarity 26.8%; Pred. No. 0.00023;
Matches 41; Conservative 19; Mismatches 36; Indels 57; Gaps 7;

QY 7 IQNISNNKTERSTAPL-----NTQISALRNDPKPLPQQPPAPANODQSSNTRLQPT 59
DB 25 VQAFCSDSKFGTSLNGLFTTNAQLSSLQPPCKPVP----- 62
QY 60 PPIAPAPKAPPPRLDRESPGVGNKLIIPSVGSPASSTPLPRDGTGPNSTPNNAVT 119

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/788,322
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-64383-1/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-819-013-1

Query Match 13.3%; Score 137; DB 2; Length 456;
Best Local Similarity 25.9%; Pred. No. 0.00083;
Matches 51; Conservative 26; Mismatches 58; Indels 62; Gaps 10;

QY 10 ISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANODQSSNTRLQPTPPIPAPAPKP 69
DB 120 IDNRSSQRHSPFFS-----KTLPSKPSWP-----SEKARLTSLPALTALQXP 162
QY 70 AAPRP---LDRESPGV-----ENKLIPSVGS---PASSTPLPPDGTGPNSTPNNAVT 117
DB 163 QVPPKPKGLLDEADYVVFVDNDENIHTTESSPPPEKAPWNRSTKNS-----T 216
QY 118 PVS-----QGSNSSADPKAPPP-----PPVSSGGPPTL-----GPNGLGLSEQ-- 157
DB 217 PASPGTASGRNSGAWETKSPPAAPSPLPRAGKPTTLPKTPVASQSNASSVCSEKPI 276
QY 158 -----LEHRSLOT 167
DB 277 PAERHGRSSHRQEAQVS 293

RESULT 11
US-09-080-897-2
; Sequence 2, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan B.
; APPLICANT: Welch, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-080-897-2

Query Match 13.3%; Score 136.5; DB 2; Length 1248;
Best Local Similarity 32.3%; Pred. No. 0.0029;
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQQ-----PPAPANODQSSNTRLQPTPPIPAPAPKPA-----APRDLRES----- 80
DB 578 PLPGDSGTIIIPPPAPGD-----STTPPPPPPPPPPPPLPGCTAISPPPLSGDATIPPP 633
QY 81 -PGVENKLIIPSVGSPASSTPLPPDGTGPNSTPNNAVTVPVSGSNSSSADPKAPPPPV 139
DB 634 PPLPEGVGIPSPSLPGTAIPP-----PPLPGSARIPPPPPPLPGSAGIPPPPPPLPGE 689
QY 140 SGEPTLGENPDG 152
DB 690 AGMPPPPPPLPGG 702

RESULT 12
US-09-323-735-2
; Sequence 2, Application US/09123735
; Patent No. 6197932
```

```

; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welch, Fari L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/323,735
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/080,897
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-323-735-2

```

```

Query Match 13.3%; Score 136.5; DB 3; Length 1248;
Best Local Similarity 32.3%; Pred. No. 0.0029;
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQO-----PPAPNQDQSSQNTLQPTTPIPAKPA-----APRPLDRES----- 80
Db 578 PLPGDSGTIIPPPAGD---STTPPPPPPPPPPLPGGTALSPPLSGDATIPPP 633
QY 81 -PGVENKLIPSVGSPASSTPLPDGTGNTSTNNRAVTVSVGSSNSADPKAPPPPPVS 139
Db 634 PLPEGVGIPSPSSLPGGTAIPP-----PPLPGSARIPPPPPPLPGSAGIRPPPPPLFGE 689
QY 140 SGEPTLGENPDG 152
Db 690 AGMPPPPPPLPG 702

```

```

RESULT 13
US-08-818-112-53
; Sequence 53, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-53

```

```

Query Match 13.1%; Score 135; DB 3; Length 332;
Best Local Similarity 27.6%; Pred. No. 0.0084;
Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;

QY 6 HIQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANODQSSQNTLQPTTPIPA 65
Db 7 HMQVDNLTNRKGLAALAIAAAS-ASLTVAVPATANADP-----EPAPPVPTT 57
QY 66 APKP-----APPRPLDRESGCVENKLIPLSVGSPASSTPLPDGTGNTSTNNRAVTVSVQ 121
Db 58 AASPSTAAAPAP-----ATPVAPPPPPAAANTEN-----AQSGDP 93
QY 122 GSNSSADPKAPPPPPVSSGEP-PTLGENPDG 152
Db 94 NAAPPPADPNAPPPPPVIAPNAPQPVRIIDNPVG 125

```

```

RESULT 14
US-08-818-111-53
; Sequence 53, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:27:44 ; Search time 52.0856 Seconds  
(without alignments)  
679.968 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028

Sequence: 1 TIVSFHQNISNNKTERSTA.....LPDEKFTGAQSGPQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028	100.0	1426	15	US-10-322-579-15
2	178	17.3	35	15	Sequence 15, Appli
3	167.5	16.3	707	11	US-10-322-579-5
4	162.5	15.8	574	12	Sequence 278, Appl
5	162.5	15.8	574	12	US-10-168-097A-76
6	158	15.4	802	9	US-10-239-431A-38
7	154.5	15.0	731	14	US-09-823-240-2
8	152	14.8	509	12	US-10-086-464-17
9	151.5	14.7	503	14	US-10-021-660-86
10	150.5	14.6	533	11	US-10-078-547-2
11	150.5	14.6	533	15	US-09-922-226-2
12	150.5	14.6	800	12	US-10-098-184-2
13	150	14.6	5179	9	US-10-029-386-32198
14	150	14.6	5179	10	US-09-922-217-1068
15	150	14.6	5179	14	US-09-833-263-1068
					Sequence 1068, Ap
					Sequence 1068, Ap

16	149	14.5	507	14	US-10-078-547-24	Sequence 24, Appli
17	148.5	14.4	731	14	US-10-086-464-8	Sequence 8, Appli
18	144.5	14.1	351	12	US-10-239-431A-22	Sequence 22, Appli
19	144.5	14.1	542	12	US-10-239-431A-10	Sequence 10, Appli
20	144.5	14.1	844	13	US-10-156-761-7663	Sequence 7663, Ap
21	142.5	13.9	721	14	US-10-086-464-5	Sequence 5, Appli
22	141.5	13.8	416	15	US-10-043-487-282	Sequence 282, App
23	141.5	13.8	505	12	US-10-168-097A-12	Sequence 12, Appli
24	141.5	13.8	505	12	US-10-239-431A-32	Sequence 32, Appli
25	141	13.7	666	9	US-09-791-171-70	Sequence 70, Appli
26	141	13.7	666	12	US-09-804-980-70	Sequence 70, Appli
27	141	13.7	666	16	US-10-080-170-640	Sequence 640, App
28	139.5	13.6	505	12	US-10-168-097A-56	Sequence 56, Appli
29	139.5	13.6	505	12	US-10-239-431A-36	Sequence 36, Appli
30	139	13.5	501	12	US-10-168-097A-46	Sequence 46, Appli
31	139	13.5	501	12	US-10-239-431A-35	Sequence 35, Appli
32	139	13.5	638	15	US-10-038-010-4	Sequence 4, Appli
33	138	13.4	377	15	US-10-149-819-7	Sequence 7, Appli
34	137.5	13.4	647	14	US-10-086-464-2	Sequence 2, Appli
35	137.5	13.4	647	14	US-10-086-464-4	Sequence 4, Appli
36	136.5	13.3	523	12	US-10-017-161-1982	Sequence 1982, Ap
37	136	13.2	633	12	US-10-168-097A-66	Sequence 66, Appli
38	136	13.2	633	12	US-10-239-431A-37	Sequence 37, Appli
39	135	13.1	332	12	US-10-084-843-53	Sequence 53, Appli
40	135	13.1	332	12	US-10-193-002-53	Sequence 53, Appli
41	135	13.1	332	12	US-10-098-732A-41	Sequence 41, Appli
42	134.5	13.1	200	9	US-09-925-297-700	Sequence 700, App
43	134.5	13.1	668	10	US-09-925-300-1633	Sequence 1633, Ap
44	134.5	13.1	668	12	US-10-240-154-20	Sequence 20, Appli
45	134.5	13.1	704	12	US-10-240-154-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1

US-10-322-579-15  
; Sequence 15, Application US/10322579  
; Publication NO. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHW  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 15  
; LENGTH: 1426  
; TYPE: PRT  
; ORGANISM: Human lgs/bcl9

US-10-322-579-15

Query Match 100.0%; Score 1028; DB 15; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 8.7e-59;  
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKLPQPAPANQDNSSQNTRLQPTP 60  
Db 199 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKLPQPAPANQDNSSQNTRLQPTP 258  
Qy 61 PIPAPAKPAAAPRPLDRESGKVENKLIPIVSGSPASSTPLPDGCTGPNTPNNRATPVIS 120  
Db 259 PIPAPAKPAAAPRPLDRESGKVENKLIPIVSGSPASSTPLPDGCTGPNTPNNRATPVIS 318

QY	121	QGSNSSADPKAPPPPPVSSGEPPTLGENPGLSOEQL EHRERSLQTLRDI QRM LFPDEK	180
Db	319	QGSNSSADPKAPPPPPVSSGEPPTLGENPGLSOEQL EHRERSLQTLRDI QRM LFPDEK	378
QY	181	EFTGAOSGGFPQONP	194
Db	379	EFTGAOSGGFPQONP	392

RESULT 2  
US-10-322-579-5  
; Sequence 5, Application US/10322579  
; Publication No. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 35  
; TYPE: prt  
; ORGANISM: Human lgs/bc19  
US-10-322-579-5

```

RESULT 3
US-09-919-039-278
; Sequence 278, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 278
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1
US-09-919-039-278

```

	Qy	30 RNDPKPLPQQPPAPANQDN--SSQNTRLQTPTPIPAAPKP-----AAPRPLDRESPGV 83 ::
	Dz	53 QSGKPL-PIPPPPTHQQGQPPEQQPFPPQHQRHPRHQPHQOQQQFPFPPDSKKPV 111 ::          :   :
	Qy	84 EN-KLIPSVGS--PASSTPLP-----PDGTGNSTN-NRAVTTPVSQGNSSSADP 130 :         :   :
	Dz	112 AQGPCRPAGVGSAAPPASSAEPATPTTSGAPGGSPGFTTPPPAVT-----SAPP 162 
	Qy	131 KAPPPPPYSSGEFFT 145 
	Dz	163 GAPPETPPSSGVETT 177 

```

RESULT 4
US-10-168-097A-76
; Sequence 76, Application US/10168097A
; Publication No. US20030166245A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; APPLICANT: INSTITUT CURIE
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
; FILE REFERENCE: IFB99WASP
; CURRENT APPLICATION NUMBER: US/10/168,097A
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 76
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-168-097A-76

```

RESULT 5  
US-10-239-431A-38  
; Sequence 38, Application US/10239431A  
; Publication No. US20030170726A1  
; GENERAL INFORMATION:  
; APPLICANT: FRADELIZ, JULIE  
; APPLICANT: FRIEDERICH, EVELYNE  
; APPLICANT: GOLSTEYN, ROY M.  
; APPLICANT: LOUWARD, DANIEL  
; APPLICANT: NOIREAUX, VINCENT  
; APPLICANT: SYKES, CECILE  
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING  
; TITLE OF INVENTION: TO PROTEINS OF THE Eha/VASP FAMILY, AND THEIR USES  
; FILE REFERENCE: 0508-1032  
; CURRENT APPLICATION NUMBER: US/10/239.431A  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: PCT/FR01/00843  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: FR 00/03637  
; PRIOR FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 38  
; LENGTH: 574



Db 153 APPPPRRSAGGAGALPCAGHTRRRRTTSPRSSPLSGPPGRASPRGAPPPPLLRAA 212  
QY 110 -TPNNRAVTPVSGNSGSSADKAPPPPVSSGE-----PP-----TL 146  
Db 213 PTPSPRALAP-----AAASPPPPPPPPGRCGEKKKFPFGSSGSTQTSCHAAVAAL 265  
QY 147 GBNPGLSQEQLHRSLSQTLRDITQMLFPDEKEFTCAQS 188  
Db 266 GSSPG-----RRRLPLLLRVGR-----PRSGAASG 291

## RESULT 9

US-10-078-547-2  
; Sequence 2, Application US/10078547  
; Publication No. US20020199211A1  
; GENERAL INFORMATION:  
; APPLICANT: Narayanaswamy Ramesh  
; APPLICANT: Miguel A. de la Fuente  
; APPLICANT: Ines M. Anton  
; APPLICANT: Raif S. Geha  
; TITLE OF INVENTION: WIP, A WASP-Associated Protein  
; FILE REFERENCE: 1242.1022-005  
; CURRENT APPLICATION NUMBER: US/10/078,547  
; CURRENT FILING DATE: 2002-02-19  
; PRIOR APPLICATION NUMBER: 09/599,287  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: PCT/US98/27501  
; PRIOR FILING DATE: 1998-12-22  
; PRIOR APPLICATION NUMBER: 60/101,457  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/068,533  
; PRIOR FILING DATE: 1997-12-23  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Human  
US-10-078-547-2

Query Match 14.7%; Score 151.5; DB 14; Length 503;  
Best Local Similarity 26.5%; Pred. No. 0.021;  
Matches 60; Conservative 25; Mismatches 84; Indels 57; Gaps 11;

QY 12 NNK-----TERSTAPLNTQISALRNDKPLPQQPAPANQONS-----SQNTRL-QPT 59  
Db 289 NKKPVVPTSPRSAPRPH-----LRPPPSRPPPLPPSSGNDTLPRLPQRLSLSSST 345  
QY 60 PPIPAKPAAPPRLDRESGVENKLIPSVGSASSTPLPDCTGPNSTPNNRAV--- 116  
Db 346 PPLPSFGSGPLPPPSERPPPPVRD-----PPGRSGPLPPPPVSRNGSTRALPAT 398  
QY 117 -TPVSGNSSSADPK-----APPPPVSSGEPPTLGENPDGLSQEQLHRS-- 161  
Db 399 PQLPSRGVDSRSGRPPPLPDRPSAGAPPPPPSTIRNGFQDSP---CEDENESRFX 455  
QY 162 -----ERSLQTLRDITQMLFPDEKETGA---QSGGPPQNP 194  
Db 456 FHPISDLPPPEFYVQTTSKPSKLARNESR-SGNSRNRERGPPPLPP 500

## RESULT 10

US-09-922-226-2  
; Sequence 2, Application US/09922226  
; Publication No. US20030077664A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhao, Yi  
; APPLICANT: Thacher, Scott M.  
; APPLICANT: Xiao, Jia-Hao  
; APPLICANT: Kusari, Jyotirmoy  
; APPLICANT: Chandraratna, Roshantha A.  
; TITLE OF INVENTION: Methods of Screening For Compounds That  
; TITLE OF INVENTION: Modulate Hormone Receptor Activity

; FILE REFERENCE: P-AR 4681  
; CURRENT APPLICATION NUMBER: US/09/922,226  
; CURRENT FILING DATE: 2002-01-09  
; PRIOR APPLICATION NUMBER: US 60/284,797  
; PRIOR FILING DATE: 2001-04-18  
; NUMBER OF SEQ ID NOS: 191  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 533  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-226-2

Query Match 14.6%; Score 150.5; DB 11; Length 533;  
Best Local Similarity 31.0%; Pred. No. 0.025;  
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;

QY 31 NDPKPLPQ--QPPAPANQONSQNTLQPTTPIAPA-----PKPAAPPRLDRESP 81  
Db 86 SSPNPLPQGVPPSPPG-----PPLPSTAPSLGSGGAPPPPPPLGSPFP 134  
QY 82 GVENKLIPSVGSASSTPLPDCTGPNSTPNNRAVTPVSGNSSSADPKAP----- 133  
Db 135 VISS-----SMGSPGLPPAPPFGSPVSSPQINSTVSLPGGSGPPEDVKPVLGVRGLH 190  
QY 134 -PPPPVSSG 141  
Db 191 CPPPPGGPG 199

## RESULT 11

US-10-098-184-2  
; Sequence 2, Application US/10098184  
; Publication No. US20030105333A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfahl, Magnus  
; APPLICANT: Tachdjian, Catherine  
; APPLICANT: Al-Shamma, Hussien A.  
; APPLICANT: Farjul, Andrea  
; APPLICANT: Plevnet, David P.M.  
; APPLICANT: Spruce, Lyle W.  
; APPLICANT: Fine, Richard  
; APPLICANT: Zapf, James W.  
; TITLE OF INVENTION: RXR ACTIVATING MOLECULES  
; FILE REFERENCE: 13099.0016U2  
; CURRENT APPLICATION NUMBER: US/10/098,184  
; CURRENT FILING DATE: 2002-03-08  
; PRIOR APPLICATION NUMBER: 60/274,342  
; PRIOR FILING DATE: 2001-03-08  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 533  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. US20030105333A1e =  
US-10-098-184-2

Query Match 14.6%; Score 150.5; DB 15; Length 533;  
Best Local Similarity 31.0%; Pred. No. 0.025;  
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;

QY 31 NDPKPLPQ--QPPAPANQONSQNTLQPTTPIAPA-----PKPAAPPRLDRESP 81  
Db 86 SSPNPLPQGVPPSPPG-----PPLPSTAPSLGSGGAPPPPPPLGSPFP 134  
QY 82 GVENKLIPSVGSASSTPLPDCTGPNSTPNNRAVTPVSGNSSSADPKAP----- 133  
Db 135 VISS-----SMGSPGLPPAPPFGSPVSSPQINSTVSLPGGSGPPEDVKPVLGVRGLH 190  
QY 134 -PPPPVSSG 141

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match      14.6%; Score 150; DB 9; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.27;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY 15 TERSTAPLNTQISALRNDPKLPQQPPAPANQDQSSQNTRLOTPPIAPAPKPAAP-- 72
Db 1428 TTTTPTTTPSPPIITTTTLPITTPSPPISTTTTTPPT-TTPSPPTTTPSPPTTSP 1486
QY 73 -----PRPDRSPGVENKLIPIVSGSPASSTPLPDGT-----GPNSTPNR 114
Db 1487 PTTTTTPTTTPSPMTPTTP-----PASTTLPPTTTPSPPTTTTTPPTTTPSP 1542
QY 115 AVTPVSGSNSSADPKAPP-----PPVSSGEPPT 145
Db 1543 TTTPTPTTSTTLPTTTPSPPTTTTTPPTTTPSPPT 1582

RESULT 14
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1068

Query Match      14.6%; Score 150; DB 10; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.27;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY 15 TERSTAPLNTQISALRNDPKLPQQPPAPANQDQSSQNTRLOTPPIAPAPKPAAP-- 72
Db 1428 TTTTPTTTPSPPIITTTTLPITTPSPPISTTTTTPPT-TTPSPPTTTPSPPTTSP 1486
QY 73 -----PRPDRSPGVENKLIPIVSGSPASSTPLPDGT-----GPNSTPNR 114
Db 1487 PTTTTTPTTTPSPMTPTTP-----PASTTLPPTTTPSPPTTTTTPPTTTPSP 1542
QY 115 AVTPVSGSNSSADPKAPP-----PPVSSGEPPT 145
Db 1543 TTTPTPTTSTTLPTTTPSPPTTTTTPPTTTPSPPT 1582

RESULT 15
US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Meagher, Madeleine Joy

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32198
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007663.28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
US-10-029-386-32198

Query Match      14.6%; Score 150.5; DB 12; Length 800;
Best Local Similarity 33.6%; Pred. No. 0.038;
Matches 41; Conservative 10; Mismatches 52; Indels 19; Gaps 5;

QY 33 PRLPQQPPAPANQDQSSQNTRLOTPP-----IPAPAKPAAPRPLDRSPGVENKLI 88
Db 609 PRLPQLPPLPPPSF-----LPPPPPSFSLPPLPPPPPSLPPPLPPPSLPPSP 660
QY 89 PSVSGSPASSTPLPDGTGPNSTPNRRAVTPVSGSNSSADPKAPPVSSGEPPTLGE 148
Db 661 PSLSP--SLPLP--PPSLPPLPPLPPLPPPSLPPSP-----APPSLPP 713
QY 149 NP 150
Db 714 SP 715

RESULT 13
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124

```

```

; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match      14.6%; Score 150; DB 14; Length 5179;
Best Local Similarity 25.6%; Pred.No. 0.27; 65; Indels 34; Gaps 5;
Matches 41; Conservative 20; Mismatches 20;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQPTTPIIPAPAPKPAAP-- 72
Db 1428 TTTTPPTTTPSPPIITTTTTLPTTTTSPPISTTTTTPPT-TTPSPPTTTPSPPTTTPSP 1486

QY 73 -----PRLDRESGVENKLIPSVGSPASSTPLPPDGT-----GPNSTPNNR 114
Db 1487 PTTTTPPTTTPSPPTTTPPTTP-----PASTTLPPTTTPSPPTTTPPTTTPSPPT 1542

QY 115 AVTPVSGQSGNSSADPKAPP-----PPPVSGCEPPT 145
Db 1543 TTTTPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPSPPT 1582

```

Search completed: November 13, 2003, 09:39:45  
Job time : 53.0856 secs









S49915  
 extensin-like protein - maize  
 C:Species: Zea mays (maize)  
 C>Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
 C:Accession: S49915  
 R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: Pez genes: pollen-specific genes with extensin-like domains.  
 A:Reference number: S49915  
 A:Accession: S49915  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-118 <RUB>  
 A:Cross-references: EMBL:Z34465; NID:G600117; PIDN:CAA84230.1; PID:G600118

Query Match 15.4%; Score 158.5; DB 2; Length 1188;  
 Best Local Similarity 32.0%; Pred. No. 0.057;  
 Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

QY 14 KTERSTAPLNTQISALRNDPKLP--QQPPAPANQDNSSQNTLRQP---TPPIAPAPAK 68  
 DB 961 KSPPPAPVNLPPPEVKSSPPPTVSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSS 1020  
 QY 69 P-----AAPPLDRESGVENKLIPIVSGSPASSTPLP---PDGTGNSTNNRAVTPV 119  
 DB 1021 PPPPVKSPPPAPVSSPPPVKS---PPPPAPVSSPPPVKSPPPAPVSSPPPVKSP 1077  
 QY 120 SGGSSSSADP-KAPPPP-PVSSGPPPTLGENP 150  
 DB 1078 PPAPVSSPPPVKSPPPAPVSSPPPVKSPPP 1110

RESULT 11  
 S22697  
 extensin - Volvox carteri (fragment)  
 C:Species: Volvox carteri  
 C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 21-Jul-2000  
 C:Accession: S22697; S21006  
 R:Erftl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.  
 EMBL J. 11, 2053-2062, 1992  
 A:Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca  
 A:Reference number: S22697; MUID:92289669; PMID:1600938  
 A:Accession: S22697  
 A:Molecule type: mRNA  
 A:Residues: 1-464 <HAL>  
 A:Cross-references: EMBL:X65165; NID:G21991; PIDN:CAA46283.1; PID:G21992  
 C:Keywords: glycoprotein

Query Match 15.3%; Score 157.5; DB 2; Length 464;  
 Best Local Similarity 31.3%; Pred. No. 0.026;  
 Matches 41; Conservative 11; Mismatches 64; Indels 15; Gaps 3;

QY 21 PLNTQISALRNDPKLPQQPPAPANQDNSSQNTLRQPPTIPAPAPKAPAPRPLDRES 80  
 DB 292 PPPRVSPPPPQPVSSPP 351  
 QY 81 PGVENKLIPIVSGSPASSTPLPDDGTGNSTNNRAVTPVSGSSSSADPKAPP-PPVS 139  
 DB 352 P---PRSSSPSPPPVSSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 397

QY 140 SGEPTLGENP 150  
 DB 398 PPPPATAAANP 408

RESULT 12  
 S16748  
 proline-rich protein - rape (fragment)  
 C:Species: Brassica napus (rape)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Sep-1997  
 C:Accession: S16748  
 R:Roberts, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, R.J.  
 submitted to the EMBL Data Library, August 1991

A:Description: Cloning and characterization of a proline-rich gene expressed specifi  
 A:Reference number: S16748  
 A:Accession: S16748  
 A:Molecule type: mRNA  
 A:Residues: 1-449 <ROB>  
 A:Cross-references: EMBL:X60376; NID:G22596; PID:G22597

Query Match 15.2%; Score 156; DB 2; Length 449;  
 Best Local Similarity 33.3%; Pred. No. 0.03;  
 Matches 39; Conservative 9; Mismatches 57; Indels 12; Gaps 3;

QY 33 PKPLPQQPPAPANQDNSSQNTLRQPPTIPAPAPKAPAPRPLDRESGVENKLIPIVSG 92  
 DB 2 PKPQKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 61  
 QY 93 -----SPASSSTPLPDDGTGNSTNNRAVTPVSGSSSSADPKAPPVSSGSP 144  
 DB 62 PSPKPGSPSPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 111

RESULT 13  
 T04455  
 hypothetical protein F4D11.90 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 04-Mar-2000  
 C:Accession: T04455  
 R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Hoheisel, J.; Mewes,  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: Z15360  
 A:Accession: T04455  
 A:Molecule type: DNA  
 A:Residues: 1-731 <BEV>  
 A:Cross-references: EMBL:AL022537  
 A:Experimental source: cultivar Columbia; BAC clone F4D11  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 326/1; 476/1; 505/1; 528/3; 557/2; 606/3; 660/3  
 A:Note: F4D11.90  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

Query Match 15.0%; Score 154.5; DB 2; Length 731;  
 Best Local Similarity 27.6%; Pred. No. 0.06;  
 Matches 47; Conservative 20; Mismatches 68; Indels 35; Gaps 7;

QY 15 TERSTAPLNTQISALRNDPKLPQQPPAPANQDNSSQNTLRQPPTIPAPAPK--PAAP 72  
 DB 29 TSSPPAPPLSLPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPLSPPL 88  
 QY 73 PRPLDRESGVENKLIPIVSGSPASSTPL-----PPDGTGNST-PNNRAVTPVSG- 122  
 DB 89 PPLESPSP-----SPHVSAPSGPPLPPLPAKPSPPSPSPSPSPSPSPSPSPSPSP 143  
 QY 123 -----SNSSADPKAPP-----PPVSSGPPPTLGENPDGLSQ 155  
 DB 144 PSESTPVNTASPPSP 192

RESULT 14  
 T05441  
 proline-rich protein F7K2.50 - Arabidopsis thaliana (fragment)  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 26-Aug-1999  
 C:Accession: T05441  
 R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue  
 submitted to the Protein Sequence Database, November 1998  
 A:Reference number: Z15416  
 A:Accession: T05441  
 A:Molecule type: DNA  
 A:Residues: 1-379 <BEV>  
 A:Cross-references: EMBL:AL033545  
 A:Experimental source: cultivar Columbia; BAC clone F7K2  
 C:Genetics:  
 A:Map position: 4

A;Note: F7K2.50  
C;Superfamily: hydroxyproline-rich glycoprotein

Query Match 14.8%; Score 152.5; DB 2; Length 379;  
Best Local Similarity 31.8%; Pred. No. 0.041;  
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY 33 PKELPQQPPAPANQDNSSQNTRLQPTPPI---PAPAPKPAAPRPLDRESGVENKLP 89  
DB 51 PQDFDQPTTPT-----FQFAPPANDQPPPPQSTSP--PVATTPPALPPKPLP 98

QY 90 SVGSPASSTPLRPDGTGPNSTNNRAVTPVSGNSSSADPKAPPPPPVSSGEPPTLGEN 149  
DB 99 PPLSPQTTTPPPPAITPPPPP---AITP-----PUSPPPPAIT--PPPLATT 142

QY 150 PDGLSQEQL 158  
DB 143 PPALPPKPL 151

RESULT 15  
D85257  
extensin-like protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
C;Accession: D85257  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: D85257  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-379 <STO>  
A;Cross-references: GB:NC\_001368; NID:g7269093; PIDN:CAB79202.1; GSPDB:GN00140  
C;Genetics:  
A;Gene: At4g22470  
A;Map position: 4  
C;Superfamily: hydroxyproline-rich glycoprotein

Query Match 14.8%; Score 152.5; DB 2; Length 379;  
Best Local Similarity 31.8%; Pred. No. 0.041;  
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY 33 PKELPQQPPAPANQDNSSQNTRLQPTPPI---PAPAPKPAAPRPLDRESGVENKLP 89  
DB 51 PQDFDQPTTPT-----FQFAPPANDQPPPPQSTSP--PVATTPPALPPKPLP 98

QY 90 SVGSPASSTPLRPDGTGPNSTNNRAVTPVSGNSSSADPKAPPPPPVSSGEPPTLGEN 149  
DB 99 PPLSPQTTTPPPPAITPPPPP---AITP-----PUSPPPPAIT--PPPLATT 142

QY 150 PDGLSQEQL 158  
DB 143 PPALPPKPL 151

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 17.3619 Seconds  
(without alignments)  
525.472 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028

Sequence: 1 TIVSFHQNISNNKTERSTA.....LFPDSKFTGAQGGPQQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	99.6	1426	1 BCL9_HUMAN	O00512 homo sapien
2	167.5	16.3	707	1 SPQ_HUMAN	P23246 homo sapien
3	151.5	15.7	815	1 MK07_HUMAN	Q13164 homo sapien
4	159.5	15.5	555	1 GNL1_HUMAN	Q91696 chlamydomon
5	158	15.4	802	1 ENAF_MOUSE	Q03173 mus musculus
6	156	15.2	449	1 APG_ERANK	P40603 brassica na
7	151.5	14.7	503	1 WAP_HUMAN	O43516 homo sapien
8	150.5	14.6	251	1 PRP2_HUMAN	P02812 homo sapien
9	150.5	14.6	533	1 MUC2_HUMAN	P28702 homo sapien
10	150	14.6	5179	1 MUC2_HUMAN	Q02817 homo sapien
11	149	14.5	426	1 EXLP_TOBAC	Q03211 nicotiana t
12	147	14.3	347	1 CSP_FLABA	P23093 plasmodium
13	146	14.2	534	1 APG_ARATH	P40602 arabidopsis
14	146	14.2	2167	1 SHK1_RAT	Q9wv48 rattus norv
15	144.5	14.1	339	1 CSP_PLABE	P06915 plasmodium
16	144.5	14.1	542	1 ZYX_CHICK	Q04584 gallus gall
17	144	14.0	535	1 SPKC_SVNY3	P74745 synechocyst
18	144	14.0	817	1 VRP1_YEAST	P37370 saccharomyc
19	143.5	14.0	520	1 RXRB_MOUSE	P28704 mus musculus
20	143	13.9	296	1 PRP3_MOUSE	P05143 mus musculus
21	143	13.9	1206	1 FMN1_MOUSE	Q05859 mus musculus
22	143	13.9	1468	1 FMN1_MOUSE	Q05860 mus musculus
23	142.5	13.9	276	1 PRPL_HUMAN	P10162 homo sapien
24	141.5	13.8	283	1 EXTN_SORBI	P24152 sorghum bic
25	141.5	13.8	505	1 WASL_HUMAN	O00401 homo sapien
26	141.5	13.8	620	1 EXTN_TOBAC	P13983 nicotiana t
27	140.5	13.7	582	1 MKT_HUMAN	Q95833 homo sapien
28	140	13.6	806	1 MK07_MOUSE	Q9wv88 mus musculus
29	139.5	13.6	268	1 NO20_MEDTR	P93129 medicago tr
30	139.5	13.6	505	1 WASL_BOVIN	P93107 bos taurus
31	139.5	13.6	1664	1 SLP1_CLOTM	Q06852 clostridium
32	139	13.5	501	1 WASL_RAT	O08816 rattus norv
33	138.5	13.5	2161	1 SHK1_HUMAN	Q9Y566 homo sapien

RESULT 1  
BCL9\_HUMAN  
ID BCL9\_HUMAN STANDARD; PRT; 1426 AA.  
AC O00512;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE B-cell lymphoma 9 protein (Bcl-9) (legless homolog).  
GN BCL9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX MEDLINE=39158621; PubMed=9490669;  
RA Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,  
RA Jadayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,  
RA Dyer M.J.S.;  
RT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel  
RT gene (BCL9) at chromosome 1q21.";  
RL Blood 91:1873-1881 (1998).  
RN [2]  
FUNCTION.  
MEDLINE=21952490; PubMed=11955446;  
RA Kramps T., Peter O., Brunner E., Neillen D., Froesch B., Chatterjee S.,  
RA Murone M., Zuelig S., Basler K.,  
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
RT pygopus to the nuclear beta-catenin-TCF complex.";  
RL Cell 109:47-60 (2002).  
CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.  
CC -!- SUBUNIT: Binds to beta-catenin (CTNBB1), PYGO1 and PYGO2.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,  
CC testis, ovary and small intestine, and at lower levels in spleen,  
CC colon and blood.  
CC -!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation  
CC found in a patient with precursor B-cell acute lymphoblastic  
CC leukemia (ALL). This translocation leaves the coding region  
CC intact, but may have pathogenic effects due to alterations in the  
CC expression level of BCL9. Several cases of translocations within  
CC the 3' untranslated region of BCL9 have been found in B-cell  
CC malignancies.  
CC -!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.  
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 1391.

## ALIGNMENTS

34 138.5 13.5 3164 1 TEGU\_HSV11  
35 137.5 13.4 234 1 PRPM\_HUMAN  
36 137 13.3 261 1 PRP2\_MOUSE  
37 137 13.3 279 1 Y091\_NPOV  
38 137 13.3 512 1 GAG\_SMSAV  
39 137 13.3 516 1 Y369\_TREPA  
40 136.5 13.3 458 1 RXRB\_HUMAN  
41 136.5 13.3 1248 1 DIAL\_HUMAN  
42 136.5 13.3 1409 1 AEX3\_CAEEL  
43 136.5 13.3 2142 1 BAT2\_HUMAN  
44 136 13.2 633 1 LAI7\_YEAST  
45 136 13.2 670 1 SYN1\_MOUSE

P10220 herpes simp  
P10161 homo sapien  
P05142 mus musculus  
O10341 oryza pscu  
P03330 simian sarc  
O83384 treponema p  
P49743 rattus norv  
O02626 caenorhabdi  
P48634 homo sapien  
Q12446 saccharomyc  
O88935 mus musculus



QY 84 EN--KLIPSVGS--PASSTPLP-----PDGTGNSTPN-NRAVTPVSGNSSSADP 130  
 DB 112 AQPGGAPGVGAPASSAPPATPTSGAPGSGEPTPTTPPAVT-----SAPP 162  
 QY 131 KAPPPPPVSSGEPT 145  
 DB 163 GAPPPTPPSSGVPT 177

## RESULT 3

MX07 HUMAN  
 ID MK07 HUMAN STANDARD; PRT; 815 AA.  
 AC Q13164; Q16634;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-regulated kinase 5) (ERK-5) (ERK4) (EMK1 kinase).  
 GN MAPK7 OR PRK7 OR ERK5 OR ERK4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=95279403; PubMed=7759517;  
 RA Zhou G., Bao Z.Q., Dixon J.E.;  
 RT "Components of a new human protein kinase signal transduction pathway".  
 RL J. Biol. Chem. 270:12665-12669 (1995).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95374539; PubMed=7646528;  
 RA Lee J.-D., Ulevitch R.J., Han J.;  
 RT "Primary structure of EMK1: a new mammalian map kinase.";  
 RL Biochem. Biophys. Res. Commun. 213:715-724 (1995).  
 CC -!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.  
 CC -!- ENZYME REGULATION: Activated by tyrosine and threonine phosphorylation (By similarity).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE IN LIVER.  
 CC -!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.  
 CC -!- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES. WHEN THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY ROLE, IS ABSENT.  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. MAP KINASE SUBFAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

DR EMBL; U25278; AAA81381.1; -;  
 DR EMBL; U29725; AAA82931.1; -;  
 DR EMBL; U29726; AAA82932.1; -;  
 DR EMBL; U29727; AAA82933.1; -;  
 DR PIR; B56708; B56708.  
 DR HSSP; P24941; 1HCL.  
 DR Genew; HGNC:6880; MAPK7.  
 DR MIM; 602521; -;  
 DR GO; GO:0004707; F:MAP kinase activity; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR InterPro; IPR003527; Map\_kin.

DR InterPro; IPR000719; Prot Kinase.  
 DR InterPro; IPR002290; Ser Thr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS01351; MAPK; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00111; PROTEIN KINASE-DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle; Phosphorylation.  
 KW Phosphorylation.  
 FT DOMAIN 54 346 PROTEIN KINASE.  
 FT DOMAIN 337 340 POLY-ALA.  
 FT DOMAIN 433 464 PRO-RICH 1.  
 FT DOMAIN 520 523 POLY-ARG.  
 FT DOMAIN 577 699 PRO-RICH 2.  
 FT BIND 60 68 ATP (BY SIMILARITY).  
 FT BINDING 83 83 ATP (BY SIMILARITY).  
 FT ACT\_SITE 181 181 BY SIMILARITY.  
 FT MOD\_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE) (BY SIMILARITY).  
 FT MOD\_RES 220 220 PHOSPHORYLATION (ACTIVATES THE KINASE) (BY SIMILARITY).  
 FT CONFLICT 19 32 AREGRTTPHRCILCS -> GPVKVEPAHTAASVA (IN REF. 2).  
 FT CONFLICT 609 609 L -> V (IN REF. 2).  
 SQ SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;

Query Match 15.7%; Score 161.5; DB 1; Length 815;  
 Best Local Similarity 32.1%; Pred. No. 0.069;  
 Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;

QY 17 RSTAPLNTQISALRNDKPLP-----QQPPAPANQDNSSQNTRELQPTPIP 63  
 DB 576 RPAAPALTVPAPAPATPTPTVQTSPPPPPLAQPGQPGAGSTGVPQACPPP 635  
 QY 64 APAPKPAAPRPLDRSPG-----VENKLIPSVGSPASSTP-----LPPDGT 105  
 DB 636 GPAPHTGPPGPIVPAPPQIATSTSLAAQSLVPPPLGSGSTPGVLPPYPPGLPPDA 695  
 QY 106 G--PNS-----TPNNRAVTPVSGNSSSADPKAPPPPPVSSGEPTTLG 147  
 DB 696 GGAQPSMSSEPDVNLVT--QLSKSQVEDPL----PPIVSGTPTKSG 737

## RESULT 4

GPI\_CHLRE  
 ID GPI\_CHLRE STANDARD; PRT; 555 AA.  
 AC Q9PFQ6; Q03927;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Vegetative cell wall protein gpl precursor (Hydroxyproline-rich glycoprotein 1).  
 DE Glycoprotein 1).  
 GN GPI.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21159092; PubMed=11258910;  
 RX Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J., Goodenough U.W.;  
 RA "Glycosylated polyproline II rods-with-kinks as a structural motif in plant hydroxyproline-rich glycoproteins.";  
 RL Biochemistry 40:2978-2987 (2001).  
 RN [2]  
 RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.  
 RX MEDLINE=91017504; PubMed=1699225;  
 RA Adair W.S., Apt K.E.;  
 RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins.";

```
Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
-!- FUNCTION: Major component of the outer cell wall w6 (crystalline)
layer.
-!- SUBUNIT: Associates with GP2 and GP3.
-!- PTM: N-glycosylated and O-glycosylated.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF309494; AAG45420.1;
EMBL; M58496; AAG59706.1; ALT_SEQ.
Glycosylated; Q9FPQ6;
InterPro; IPR002965; P-rich_extensin.
InterPro; IPR003882; Pistil_extensin.
PRINTS; PR01217; PRICHEXTENSIN.
PRINTS; PR01218; PSTLEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29 POTENTIAL
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSEX REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 555 AA; 54219 MW; 6A584A9045502F5 CRC64;
SQ
Query Match 15.5%; Score 159.5; DB 1; Length 555;
Best Local Similarity 32.8%; Pred. NO. 0.063;
Matches 38; Conservative 6; Mismatches 49; Indels 23; Gaps 3;
QY 33 PKPLPQPPAPANQDSSQNTFLQPTPIP-----APAPKPAAPPLDRESFGVENKLI 88
Db 179 PPSPPVPVPSAPP-----SPAPVPPSPAPPSPAPPVPPSPAPPSPAPPSP 228
QY 89 PSVSGSPASPTLPDGTGPNSTPNRAVTFVSGSSADPKAPPPPVSSGEP 144
Db 229 PPSPPSPSPAPPSPVPSPAPPSPA-----PPSPKPPAPPSPSPPPPP 275
RESULT 5
ENAH MOUSE STANDARD; PRT: 802 AA.
ID Q03173; P70430; P70431; P70432; P70433;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Enabled protein homolog (NPC derived proline-rich protein 1) (NDPP-1).
GN ENAH OR MENA OR NDPP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA MEDLINE=93041923; PubMed=1420303;
RX Sazuka T., Tomooka Y., Kathju S., Ikawa Y., Noda M., Kumar S.;
RT "Identification of a developmentally regulated gene in the mouse
RL Biochim. Biophys. Acta 1132:240-248(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4 AND 5), FUNCTION, AND SUBCELLULAR
RC LOCATION;
RX TISSUE=Brain;
RA MEDLINE=97015079; PubMed=8861907;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
control of microfilament dynamics.";
Cell 87:227-239(1996).
[3]
RN FUNCTION, AND SUBUNIT.
RP MEDLINE=9916867; PubMed=10069337;
RX Lanier L.M., Gates M.A., Witke W., Menzies A.S., Wehman A.M.,
RA Macklis J.D., Kwiatkowski D., Soriano P., Gertler F.B.;
RT "Mena is required for neurulation and commissure formation.";
RL Neuron 22:313-325(1999).
CC -!- FUNCTION: May be involved in microfilament assembly and cell
motility. Induces the formation of F-actin rich outgrowths in
fibroblasts. Required for neurulation and commissure formation.
CC -!- SUBUNIT: Binds profilin.
CC -!- SUBCELLULAR LOCATION: Localized to focal adhesions and, to a
lesser extent, leading edges and stress fibers.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=5; Synonyms=Mena++;
CC IsoId=Q03173-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q03173-2; Sequence=VSP_007255, VSP_007256;
CC Name=2; Synonyms=Mena;
CC IsoId=Q03173-3; Sequence=VSP_007259, VSP_007260;
CC Name=3; Synonyms=Mena+;
CC IsoId=Q03173-4; Sequence=VSP_007259;
CC Name=4; Synonyms=Mena++;
CC IsoId=Q03173-5; Sequence=VSP_007257, VSP_007258;
CC -!- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
CC -!- SIMILARITY: Contains 1 WH1 domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D10727; BAA01570.1;
EMBL; U72520; AAC52863.1;
EMBL; U72521; AAC52864.1;
EMBL; U72522; AAC52865.1;
EMBL; U72523; AAC52866.1;
PIR; S27200; S27200.
MGD; MGI:108360; Enah.
InterPro; IPR001960; WH1.
SMART; SM00461; WH1; 1.
Developmental protein; Neurogenesis; Alternative splicing.
DOMAIN 1 108 WH1.
FT DOMAIN 442 464 POLY-PRO.
FT DOMAIN 542 552 POLY-PRO.
FT DOMAIN 562 574 POLY-PRO.
FT DOMAIN 578 589 POLY-PRO.
FT DOMAIN 593 605 POLY-PRO.
FT VARSPLIC 1 412 Missing (in isoform 1).
FT VARSPLIC 535 631 Missing (in isoform 1).
FT VARSPLIC 117 131 Missing (in isoform 4).
FT VARSPLIC 132 135 CIRC -> VFYL (in isoform 4).
FT VARSPLIC 117 135 /FTid=VSP_007258.
FT VARSPLIC 259 500 Missing (in isoform 2).
FT SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;
```







RA Numasawa T., Koga H., Ueyama K., Maeda S., Sakou T., Harata S.,  
RA Leppert M., Inoue I.,  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Corolla A., Vergara A., Paerz G., de Miguel C., Encio I.,  
RT "Molecular cloning and characterization of the human HXRB gene and 5'  
RT flanking region.",  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Tubby B.,  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=2238257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunatathne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnarch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.",  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16893-16903(2002).  
CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS 9-CIS  
CC RETINOIC ACID (9C-RA).  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=Long;  
CC IsoId=P28702-1; Sequences=Displayed;  
CC Name=Short;  
CC IsoId=P28702-2; Sequences=Not described;  
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR2  
CC subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X63522; CAA45087.1; -;  
CC EMBL; M84820; AAA60293.1; -;  
CC EMBL; AF065396; AAC18599.1; -;  
CC EMBL; AF120161; AAD13794.1; -;  
CC EMBL; AF031228; CAA20239.1; -;  
CC EMBL; BC001167; AAH01167.1; -;  
CC PIR; S37781; S37781.  
CC PDB; 1H9U; 22-MAY-02.  
CC TRANSFAC; T01334; -;  
CC Genew; HGNC:10478; RXRB.  
CC MIM; 180246; -;  
CC GO; GO:0004886; F:retinoid-X receptor activity; TAS.  
CC GO; GO:0003713; F:transcription co-activator activity; TAS.

DR InterPro; IPR000536; Hormone\_rec\_lig.  
DR InterPro; IPR001723; Steroid\_receptor.  
DR InterPro; IPR001638; Znf\_C4steroid.  
DR Pfam; PF00104; hormone\_rec; 1.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR00398; STROHORMONER.  
DR PRINTS; PR00047; STROIDFINGER.  
DR ProDom; PD000035; Znf\_C4steroid; 1.  
DR SMART; SM00430; HOU1; 1.  
DR SMART; SM00399; Znf\_C4; 1.  
DR PROSITE; PS0031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger; Multigene family; Alternative splicing; 3D-structure.  
FT DOMAIN 1 204 MODULATING (BY SIMILARITY).  
FT DNA\_BIND 205 270 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 205 225 C4-TYPE.  
FT ZN\_FING 241 265 HINGE.  
FT DOMAIN 271 330 HINGE.  
FT DOMAIN 331 533 LIGAND-BINDING (BY SIMILARITY).  
FT DOMAIN 61 198 PRO-RICH.  
FT CONFLICT 112 112 S -> T (IN REF. 2).  
SQ SEQUENCE 533 AA; 56921 MW; D0069FE93AC16A04 CRC64;  
Query Match 14.6%; Score 150.5; DB 1; Length 533;  
Best Local Similarity 31.0%; Pred. No. 0.17;  
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;  
QY 31 NDPKLPQ--QPPAPANQDNSSQNRLOFTPTPIAPA-----PKPAAPRPLDRESP 81  
DB 86 SSNPLPQGVPPSPFG-----PPLPSTAPSLGSGCAPPPPPPLGSPFP 134  
QY 82 GVENKLIPVSGSPASSTPLPDGTGPNSTNNRAVTFVSGNSSSADPKAP----- 133  
DB 135 VISS-----SMGSPQLPAPPAPGPGSVSPQINSTVSLPGSGGPPEDVKPVLGVRGLH 190  
QY 134 -PPPVSSG 141  
DB 191 CPPPGGPG 199  
RESULT 10  
MUC2\_HUMAN  
ID MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
AC Q02817; Q14878;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DB Mucin 2 precursor (Intestinal mucin 2).  
GN MUC2 OR SMUC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Intestine;  
RX MEDLINE=94132002; PubMed=8300571;  
RX Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.  
RT Identification of the amino terminus and overall sequence similarity  
RT to prepro-von Willebrand factor.",  
RL J. Biol. Chem. 269:2440-2446(1994).  
RN [2]  
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
RC TISSUE=Colon;  
RX MEDLINE=93016075; PubMed=1400449;  
RX Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
RX Kim Y.S.;  
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
RT both upstream and downstream of its central repetitive region.",  
RL J. Biol. Chem. 267:21375-21383(1992).  
RN [3]  
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.

RX MEDLINE=91358717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.M., Kim Y.S.;  
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 RT and polymorphism.";  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND  
 CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A  
 CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS  
 CC AGENTS AT MUCOSAL SURFACES.  
 CC -!- SUBUNIT: MULTIMERIC.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
 CC BRONCHUS, CERVIX AND GALL BLADDER.  
 CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
 CC VARIES AMONG DIFFERENT ALLELES.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC -!- SIMILARITY: Contains 1 TIL (trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 2 VWFC domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L2198; AAB95295.1; -;  
 DR EMBL; M74027; AAA59875.1; -;  
 DR EMBL; M94131; AAA59163.1; -;  
 DR EMBL; M94132; AAA59164.1; -;  
 DR PIR; A49963; A43932.  
 DR Genew; HGNC:7512; MUC2.  
 DR MIM; 158370; -;  
 DR GO; GO:0005803; C:secretory vesicle; TAS.  
 DR InterPro; IPR006208; Cys knot.  
 DR InterPro; IPR006207; Cys knot\_C.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR002919; TIL\_Cysrich.  
 DR InterPro; IPR001007; VWFC.  
 DR InterPro; IPR001846; VWFC\_D.  
 DR Pfam; PF00007; Cys\_knot; 1.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; vwc; 1.  
 DR Pfam; PF00094; vwd; 4.  
 DR SMART; SM00214; VWC; 2.  
 DR SMART; SM00216; VWD; 4.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01208; VWFC\_1; 2.  
 DR PROSITE; PS0184; VWFC\_2; 2.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 5179 MUCIN 2.  
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
 FT REPEAT 1401 1416 1.  
 FT REPEAT 1417 1432 2.  
 FT REPEAT 1433 1448 3.  
 FT REPEAT 1449 1464 4.  
 FT REPEAT 1465 1471 5.  
 FT REPEAT 1472 1478 6.  
 FT REPEAT 1479 1494 7A.  
 FT REPEAT 1495 1517 7B.  
 FT REPEAT 1518 1533 8A.  
 FT REPEAT 1534 1556 8B.  
 FT REPEAT 1557 1572 9A.  
 FT REPEAT 1573 1596 9B.  
 FT REPEAT 1597 1612 10A.  
 FT REPEAT 1613 1635 10B.

FT REPEAT 1651 1651  
 FT REPEAT 1652 1675  
 FT REPEAT 1676 1683  
 FT REPEAT 1684 1699  
 FT REPEAT 1700 1715  
 FT REPEAT 1716 1731  
 FT REPEAT 1732 1747  
 FT DOMAIN 4815 4886  
 FT DOMAIN 4924 4991  
 FT DOMAIN 5075 5160  
 FT DISULFID 5075 5122  
 FT DISULFID 5089 5136  
 FT DISULFID 5098 5152  
 FT DISULFID 5102 5154  
 FT DISULFID 5102 5159  
 FT CARBOHYD 163 163  
 FT CARBOHYD 423 423  
 FT CARBOHYD 670 670  
 FT CARBOHYD 770 770  
 FT CARBOHYD 894 894  
 FT CARBOHYD 1139 1139  
 FT CARBOHYD 1154 1154  
 FT CARBOHYD 1215 1215  
 FT CARBOHYD 1230 1230  
 FT CARBOHYD 1246 1246  
 FT CARBOHYD 1787 1787  
 FT CARBOHYD 1820 1820  
 FT CARBOHYD 4339 4339  
 FT CARBOHYD 4351 4351  
 FT CARBOHYD 4362 4362  
 FT CARBOHYD 4373 4373  
 FT CARBOHYD 4422 4422  
 FT CARBOHYD 4438 4438  
 FT CARBOHYD 4502 4502  
 FT CARBOHYD 4616 4616  
 FT CARBOHYD 4627 4627  
 FT CARBOHYD 4752 4752  
 FT CARBOHYD 4787 4787  
 FT CARBOHYD 4881 4881  
 FT CARBOHYD 4888 4888  
 FT CARBOHYD 4955 4955  
 FT CARBOHYD 4970 4970  
 FT CARBOHYD 5019 5019  
 FT CARBOHYD 5038 5038  
 FT CARBOHYD 5069 5069  
 FT CONFLICT 1351 1351  
 FT CONFLICT 1412 1412  
 FT CONFLICT 1449 1449  
 FT CONFLICT 1504 1504  
 FT CONFLICT 4192 4192  
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;  
 Query Match 14.6%; Score 150; DB 1; Length 5179;  
 Best Local Similarity 25.6%; Pred. No. 1.3;  
 Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;  
 QY 15 TERSTAPLNTQISALRNDPKLPQQPPAPANQDNSSONTROLQTPPIAPAPKPAAP-- 72  
 Db 1428 TTTTPPTTTPSPITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSP 1486  
 QY 73 -----PRLDRESGVENKLIPSGVSGPASSTPLPPDGT-----GPNSTPNR 114  
 Db 1487 PTTTTTPPTTTPSPMTTTP-----PASTTTPPTTTPSPPTTTPPTTTPPTTTPSP 1542  
 QY 115 AVTPVSGSNSSSADKAPP-----PPVSSSGSPPT 145  
 Db 1543 TTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSPPT 1582  
 RESULT 11  
 EXLP TOBAC  
 ID EXLP TOBAC STANDARD; PRT; 426 AA.  
 AC Q03211;

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-JUL-1999 (Rel. 38, Last annotation update)  
 DE Pistil-specific extensin-like protein precursor (PELP).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Petit Havana; TISSUE=Pistil;  
 RX MEDLINE=93005740; PubMed=1392607;  
 RA Goldman M.H., Pezrotti M., Seurinck J., Mariani C.;  
 RT "Developmental expression of tobacco pistil-specific genes encoding  
 novel extensin-like proteins.";  
 RL Plant Cell 4:1041-1051(1992).  
 CC -!- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER  
 PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING  
 FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE  
 AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER  
 POLLINATION.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z14019; CAA78337.1; -  
 DR PIR; JQ1696; JQ1696.  
 DR InterPro; IPR006041; Ole\_el\_ext.  
 DR InterPro; IPR002965; P-rich\_extensin.  
 DR InterPro; IPR003882; Pistil\_extensin.  
 DR Pfam; PF01190; Pollen\_Ole\_e\_I; 1.  
 DR PRINTS; PR01217; PRICHEXTEN.  
 DR PRINTS; PR01218; PSTLEXTENSIN.  
 KW Structural protein; Signal; Repeat; Glycoprotein.  
 FT SIGNAL 1 23  
 FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.  
 FT DOMAIN 69 182 4 X 5 AA REPEATS OF S-P(4).  
 FT REPEAT 69 73 1.  
 FT REPEAT 76 80 2.  
 FT REPEAT 83 87 3.  
 FT REPEAT 178 182 4.  
 FT CARBOHYD 310 310 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 426 AA; 44278 MW; 51A495CC94017812 CRC64;  
 Query Match 14.5%; Score 149; DB 1; Length 426;  
 Best Local Similarity 29.9%; Pred. No. 0.17;  
 Matches 40; Conservative 14; Mismatches 54; Indels 26; Gaps 5;  
 QY 31 NDPKLPQPPAPANQDONS--SQNTRLOTPPIIPAPKPAAPRPLDRSPGVENKLI 88  
 Db 140 SSPSLVLPKPPPPPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 191  
 QY 89 PSVGSFASSTLPDPGT-GPNSPTNNRAVTPVSGNSNSSADPKAPPPPPVVS----- 139  
 Db 192 ---SPAKQPPPPPPVKAAPSATQPTKQPPPPPPRAKKSPLLPPPPVAYPVMTFS 247  
 QY 140 ---SCEPPTLGENP 150  
 Db 248 PSPAAEPPIIAPFP 261  
 RESULT 12  
 CSP\_PLABA STANDARD; PRT; 347 AA.  
 AC F23093;  
 DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Circumsporozoite protein precursor (CS).  
 OS Plasmodium berghei (strain Anka).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90221834; PubMed=2183186;  
 RA Lockyer M.J., Davies C.S., Suhrbier A., Sinden R.E.;  
 RT "Nucleotide sequence of the Plasmodium berghei circumsporozoite  
 protein gene from the ANKA clone 2.34L.";  
 RL Nucleic Acids Res. 18:376-376(1990).  
 CC -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE  
 MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 VERTEBRATE HOST).  
 CC -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -!- SIMILARITY: Contains 1 TSP type-1 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X17606; CAA35608.1; -  
 DR PIR; S07873; OZ2QBK.  
 DR InterPro; IPR003067; Circmsprzoite.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR PRINTS; PR01303; CRCMSPRZOITE.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS00092; TSP1; 1.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 347 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 93 204 13 X 8 AA REPEATS.  
 FT DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.  
 FT DOMAIN 274 325 TSP TYPE-1.  
 SQ SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;  
 Query Match 14.3%; Score 147; DB 1; Length 347;  
 Best Local Similarity 29.9%; Pred. No. 0.18;  
 Matches 47; Conservative 10; Mismatches 68; Indels 32; Gaps 8;  
 QY 8 QNISNNKTERSTAPLNTQISALRNDPKPLQPPAPANQDONSQNTLRLOTPP---IP 63  
 Db 76 KNEKKNKIER-----NNKLK--QPPPPPNNDPPPPPNND-----PPPPNNDPP 118  
 QY 64 APAPKPAAPRPLDRSPGVENKLIPIVSGSPA---SSTPLPDPGTGP-----NSTPNNA 115  
 Db 119 PPENNDPPPPPNNDPPPPPNNDPPPPPNNDPPPPPNNDPPPPPNNDPPPPPNNDPP 178  
 QY 116 VTPVSGNSNSSADPKA--PPPPPVSSGEPTLGENP 150  
 Db 179 NDPPPPPNNDPPPPPNNDPPPPPNNDPPPPPNNDPPPPPNNDPPPPPNNDPP 214  
 RESULT 13  
 APG\_ARATH STANDARD; PRT; 534 AA.  
 AC P40602; Q93214; Q9LNT8;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Anter-specific proline-rich protein APG precursor.  
 GN APG OR ATIG20130 OR T20H2.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).



RX MEDLINE-20020275; PubMed-10551867;  
 RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;  
 RT "Somatostatin receptor interacting protein defines a novel family of  
 RT multidomain proteins present in human and rodent brain.";  
 RL J. Biol. Chem. 274:32997-33001(1999).  
 RN [6]  
 RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.  
 RX MEDLINE-99360851; PubMed-10433269;  
 RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,  
 RA Doan A., Akalu V.K., Lanahan A.A., Sheng M., Worley P.F.;  
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of  
 RT postsynaptic density proteins.";  
 RL Neuron 23:583-592(1999).  
 RN [7]  
 RP INTERACTION WITH SPTAN1.  
 RX MEDLINE-21523912; PubMed-11509555;  
 RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,  
 RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;  
 RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the  
 RT multidomain Shank protein family interact with the cytoskeletal  
 RT protein alpha-Fodrin.";  
 RL J. Biol. Chem. 276:40104-40112(2001).  
 RN [8]  
 RP FUNCTION.  
 RX MEDLINE-21389514; PubMed-11498055;  
 RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;  
 RT "Regulation of dendritic spine morphology and synaptic function by  
 RT Shank and Homer.";  
 RL Neuron 31:115-130(2001).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE-20267867; PubMed-10806096;  
 RA Sheng M., Kim E.;  
 RT "The Shank family of scaffold proteins.";  
 RL J. Cell Sci. 113:1851-1856(2000).  
 CC -1- FUNCTION. Seems to be an adapter protein in the postsynaptic  
 CC density (PSD) of excitatory synapses that interconnects receptors  
 CC of the postsynaptic membrane including NMDA-type and metabotropic  
 CC glutamate receptors, and the actin-based cytoskeleton. May play a  
 CC role in the structural and functional organization of the  
 CC dendritic spine and synaptic junction. Overexpression promotes  
 CC maturation of dendritic spines and the enlargement of spine heads  
 CC via its ability to recruit Homer to postsynaptic sites, and  
 CC enhances presynaptic function.  
 CC -1- SUBUNIT. May homomultimerize via its SAM domain. Interacts with  
 CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with  
 CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via  
 CC the PDZ domain (By similarity).  
 CC -1- SUBCELLULAR LOCATION. Cytoplasmic; postsynaptic density of  
 CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.  
 CC -1- ALTERNATIVE PRODUCTS.  
 CC Event=Alternative splicing; Named isoforms=5;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q9WV48-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9WV48-2; Sequence=VSP\_006072, VSP\_006073;  
 CC Name=3;  
 CC IsoId=Q9WV48-3; Sequence=VSP\_006074;  
 CC Name=4; Synonyms=A;  
 CC IsoId=Q9WV48-4; Sequence=VSP\_006075;  
 CC Name=5;  
 CC IsoId=Q9WV48-5; Sequence=VSP\_006076, VSP\_006077;  
 CC -1- TISSUE SPECIFICITY. Expressed only in brain (neuropil of cortex,  
 CC CA1 region hippocampus and molecular layer of cerebellum).  
 CC -1- DEVELOPMENTAL STAGE. Expression increases from low levels at birth  
 CC to high levels at 3-4 weeks before dropping slightly in adulthood.  
 CC Expressed in the cortex and the molecular layer of the cerebellum  
 CC at postnatal day 7. Isoform 2 expression does not change during  
 CC development of both cortex and cerebellum. Isoform 4 expression  
 CC decreases significantly during development of cortex but not  
 CC cerebellum.  
 CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.

CC -1- SIMILARITY: Contains 7 ANK repeats.  
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF102855; AAD04569.2;  
 DR EMBL: AF121951; AAD29417.1; ALT\_INIT.  
 DR EMBL: AF159046; AAD42975.1;  
 DR EMBL: AF141904; AAF02498.1; ALT\_INIT.  
 DR HSP: P00519; IABL.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001660; SAM.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00023; ank; 7.  
 DR Pfam: PF00595; PDZ; 1.  
 DR Pfam: PF00536; SAM; 1.  
 DR Pfam: PF00018; SH3; 1.  
 DR ProDom: PD000066; SH3; 1.  
 DR SMART: SM00248; ANK; 6.  
 DR SMART: SM00228; PDZ; 1.  
 DR SMART: SM00454; SAM; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50088; ANK\_REPEAT; 3.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS50106; PDZ; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 DR PROSITE: PS50105; SAM\_DOMAIN; 1.  
 DR ANK repeat; SH3 domain; Repeat; Alternative splicing.  
 FT REPEAT 195 210 ANK 1.  
 FT REPEAT 212 245 ANK 2.  
 FT REPEAT 246 278 ANK 3.  
 FT REPEAT 279 312 ANK 4.  
 FT REPEAT 313 345 ANK 5.  
 FT REPEAT 346 378 ANK 6.  
 FT REPEAT 379 395 ANK 7.  
 FT DOMAIN 554 613 SH3.  
 FT DOMAIN 2104 2167 SAM.  
 FT DOMAIN 929 932 POLY-PRO.  
 FT DOMAIN 1010 1015 POLY-HIS.  
 FT DOMAIN 1022 1027 POLY-HIS.  
 FT DOMAIN 1194 1199 POLY-GLY.  
 FT DOMAIN 1850 1860 POLY-PRO.  
 FT VARSPLIC 1 613 Missing (in isoform 2).  
 FT VARSPLIC 615 654 /FTId=VSP\_006072.  
 FT VARSPLIC 615 654 SOGROESRSRDKAKRLFRHVTVGSYDFAPSLIDGSG  
 FT VARSPLIC 615 654 -> MALSAVGSGPGGALPQPPALSSSPALCPRRRSVMY  
 FT VARSPLIC 615 654 IV (in isoform 2).  
 FT VARSPLIC 615 654 /FTId=VSP\_006073.  
 FT VARSPLIC 615 654 Missing (in isoform 3).  
 FT VARSPLIC 615 654 /FTId=VSP\_006074.  
 FT VARSPLIC 615 654 Missing (in isoform 4).  
 FT VARSPLIC 615 654 /FTId=VSP\_006075.  
 FT VARSPLIC 615 654 LSPDSOTSLSKPS -> QYRIWKSSDFGDF (in  
 FT VARSPLIC 615 654 isoform 5).  
 FT VARSPLIC 615 654 /FTId=VSP\_006076.  
 FT VARSPLIC 615 654 Missing (in isoform 5).  
 FT VARSPLIC 615 654 /FTId=VSP\_006077.  
 FT CONFLICT 1141 1141 S -> T (IN REF. 1).  
 FT CONFLICT 1174 1174 S -> N (IN REF. 2).  
 FT CONFLICT 1246 1246 R -> K (IN REF. 1).  
 FT CONFLICT 1323 1323 A -> T (IN REF. 1).  
 FT CONFLICT 1331 1331 S -> D (IN REF. 1).  
 FT CONFLICT 1726 1726 S -> N (IN REF. 2).

SQ SEQUENCE 2167 AA; 226333 MW; 3F478B5A7B18BA86 CRC64;  
 Query Match 14.2%; Score 146; DB 1; Length 2167;  
 Best Local Similarity 23.8%; Pred. No. 0.96;  
 Matches 56; Conservative 35; Mismatches 82; Indels 62; Gaps 11;  
 QY 6 HIQNISNNKT-ERSTAPLNTQISALRN-----KLPQOPAPANQDQSSQNTRLQ 57  
 DB 1497 HVRFLEN--QARPAPAGTRGSGTGGVPPPPRRVLPSTSPRGNEENG-----L 1548  
 QY 58 PTEPTAPAP-----KPAAP-----RLDRSPGVENKLIPIVSGSPASS 97  
 DB 1549 PLLVLPPAPSDVDGELFAELPLPPLEFNSPEKESPLTPGPHPL-PDPPSPATP 1607  
 QY 98 TPLPP-----DGGPNSTNNRAVTPVSGSSSSADPKA--PPPPVSSGEPPTL 146  
 DB 1608 LPAAPPPAAAPPLDSTASSLTSYDSEVALTQGAAPAGDPFAPGPPAPAPAPAP 1667  
 QY 147 GENPD-----GLSQEQLHRRS---LQTLRDIQRM--LFPDKEFTGAQSGG 189  
 DB 1668 QFGPDPPPTDGGI--EEVDSRSSSDHPLETISSTLSLSAEGGNTGGVAGG 1720

RESULT 15

CSP PLABE STANDARD; PRT; 339 AA.  
 ID CSP PLABE AC P06315;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Circumsporozoite protein precursor (CS).  
 OS Plasmodium berghei.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5821;  
 RN 11)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87089740; PubMed=2432395;  
 RA Eichinger D.J., Arnot D.E., Tam J.P., Nussenzweig V., Enea V.;  
 RT "Circumsporozoite protein of Plasmodium berghei: gene cloning and  
 identification of the immunodominant epitopes.";  
 RL Mol. Cell. Biol. 6:3965-3972(1986).  
 CC -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE  
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -!- SIMILARITY: Contains 1 TSP type-1 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M14135; AAA29577.1; -;  
 DR PIR; A44948; OZQWB.  
 DR InterPro; IPR003067; Crcmsprzoite.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR PRINTS; PR01303; CRCMSPRZOITE.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 23 PROBABLE.  
 FT CHAIN 24 339 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 93 196 13 X 8 AA TANDEM REPEATS.  
 FT DOMAIN 206 238 16 X 2 AA TANDEM REPEATS OF P-Q.  
 FT DOMAIN 266 317 TSP TYPE-1.  
 SQ SEQUENCE 339 AA; 37138 MW; E8068A6D11D9551B CRC64;

Query Match 14.1%; Score 144.5; DB 1; Length 339;  
 Best Local Similarity 29.1%; Pred. No. 0.24;  
 Matches 51; Conservative 13; Mismatches 66; Indels 45; Gaps 11;  
 QY 5 PHIQNISNNKT-ERSTAPLNTQISALRN-----DPKLPQOPAPAPA 44  
 DB 48 YHVLNSKNGKIYNNRT--VNRLLPMLRRKKKQKKEIKERNKLLKQPPPPNPNPPN 105  
 QY 45 NQDQSSQNTLQTPPIIP--APAPKPAAPRP--LDRESGVENKLIPIVSGSPA---SS 97  
 DB 106 PND-----PPPPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 155  
 QY 98 TPLPPDGTGPNSTNNRAVTPVSGSSSSADPKA--PPPPVSSGEPPTLGENP 150  
 DB 156 DPAPFNANDP-APPN--ANDAPPNAN--DPAPFNANDP-PPPNPNPNPNPNPNPN 206

Search completed: November 13, 2003, 09:25:55  
 Job time : 19.3619 secs

Result No.	Query	Score	Query		DB	ID	Description
			Match	Length			
1	177.5	17.3	473	10	Q39620	Q39620 chlamydomon	
2	168.5	16.4	876	3	Q39479	Q39479 neurospora	
3	167.5	16.3	1269	10	Q8WSK6	Q8WSK6 oryza sativ	
4	163	15.9	1307	10	Q9LVN1	Q9LVN1 arabidopsi	
5	162.5	15.8	574	3	Q36027	Q36027 schizosacch	
6	162.5	15.8	1011	3	Q39F944	Q39F944 pneumocysti	
7	162.5	15.8	1151	13	Q57580	Q57580 gallus gall	
8	161.5	15.7	816	4	Q96G51	Q96G51 homo sapien	
9	160.5	15.6	2321	12	Q9DGT6	Q9DGT6 turkey herp	
10	160	15.6	488	16	Q8DGU1	Q8DGU1 synechococc	
11	160	15.6	839	16	Q9RX57	Q9RX57 deinoxococ	
12	158.5	15.4	351	10	Q39492	Q39492 chlamydomon	
13	158.5	15.4	420	5	Q3VZC2	Q3VZC2 drosophila	
14	158.5	15.4	1188	10	Q41805	Q41805 zea mays (m	
15	158	15.4	802	11	P70433	P70433 mus musculu	
16	157.5	15.3	464	10	Q41645	Q41645 volvox cart	





Best Local Similarity 29.9%; Pred. No. 0.00054;  
Matches 46; Conservative 18; Mismatches 44; Indels 46; Gaps 7;  
QY 9 NISNN--KTERSTAPLNTQISALRNDPKP-LPQOPPPAPANODQNSQNTRLQPTPPAP 65  
Db 662 NVASLWGQAPSPPPIS-----NSDKKALPRPPPPPPPPPPPPPPHSTVTKVPPPP----- 711  
QY 66 APKPAAPRPRLDRSPGVENKLIPIVSGSPASSTPLPDGTCGNSPTNNRAVTPVSGSNGS 125  
Db 712 -PAPPAPTPIVH-----TSSPPPPPPPPPPPPPP-----TQSNQISA 749  
QY 126 SSADPKASP-----PPVSSGSEPTLGE 148  
Db 750 MKSSPPAPPAPRLPETHSASPPPTATPPPLGQ 783

RESULT 5  
Q36027 PRELIMINARY; PRT; 574 AA.  
AC Q36027; 01-JUN-1998 (TREMELrel. 06, Created)  
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Wiskott-aldrich syndrome protein homolog 1.  
GN WSP1 OR SPAC4F10.15C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JS21;  
RA Zankel T.C.; Ow D.W.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Connor R.; Church C.M.; Barrell B.G.; Rajandream M.A.; Wood V.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: TO YEAST LAS17.  
DR EMBL; AF038575; AAB92587.1; -.  
DR EMBL; Z98980; CAB11718.1; -.  
DR GeneDB; SPombe; SPAC4F10.15C; -.  
DR InterPro; IPR000697; EVH1.  
DR InterPro; IPR002965; P\_rich\_extensn.  
DR Pfam; PF00468; WH1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00461; WH1; 1.  
FT DOMAIN 311 317 POLY-PRO.  
FT DOMAIN 337 343 POLY-PRO.  
FT DOMAIN 361 366 POLY-PRO.  
FT VARIANT 248 248 L -> V (IN STRAIN JS21).  
SQ SEQUENCE 574 AA; 59605 MW; C6B5EFC6A02F0E4 CRC64;

Query Match 15.8%; Score 162.5; DB 3; Length 574;  
Best Local Similarity 30.9%; Pred. No. 0.00027;  
Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
QY 7 IQNIGNKTERSTPA-----PLNTQISA---LRNDPKPLPQOPPPAPANODQ---NSSQNT 54  
Db 275 IAFVSNVPAINSTSKPLPLPPPSRSVSAALANKKPPPPPPPPPSRRNRKPKPIGNGSNGS 334  
QY 55 RLQPTPP-----IPAPKPAAPRPLDRSPGVENKLIPIVSGSPA-SSTPLPDG 104  
Db 335 SLPPPPPPPSNAAGSIPLPPQGRSAPPSPRSPSTGKQPPPLSSRAVSNPPAPPPA 394  
QY 105 TGNSTENNRAVTPVSGSNGSADPKAPPPVSSGSEPTL 146  
Db 395 IPGRSAP---ALPFLGNASTST--PPVTPPELPPSPAPSL 431

RESULT 6  
Q9P944 PRELIMINARY; PRT; 1011 AA.  
AC Q9P944;  
DT 01-OCT-2000 (TREMELrel. 15, Created)  
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Kexin-like protease KEX1.  
GN KEX1.  
OS Pneumocystis carinii f. sp. muris.  
OC Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes; Pneumocystidaceae;  
OX NCBI\_TaxID=42066;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20184731; PubMed=10721706;  
RA Lee L.H.; Gigliotti F.; Wright T.W.; Simpson-Haidaris P.J.;  
RA Weinberg G.A.; Haidaris C.G.;  
RT "Molecular characterization of KEX1, a kexin-like protease in mouse  
Pneumocystis carinii";  
RL Gene 242:141-150(2000).  
CC -I- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
DR EMBL; AF093132; AAF32493.1; -.  
DR MEROPS; S08.011; -.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR006970; PT.  
DR InterPro; IPR002884; P domain.  
DR InterPro; IPR002965; P\_rich\_extensn.  
DR Pfam; PF01483; P; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF04886; PT; 3.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR PRINTS; PR00723; SUBTILISIN.  
DR ProDom; PD000717; P domain; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
KW Protease.  
SQ SEQUENCE 1011 AA; 112021 MW; FBE472C8F65864E8 CRC64;

Query Match 15.8%; Score 162.5; DB 3; Length 1011;  
Best Local Similarity 27.8%; Pred. No. 0.00046;  
Matches 49; Conservative 21; Mismatches 65; Indels 41; Gaps 7;  
QY 18 STAPLNTQISALRNDPKPLPQOPPPAPANQ---QNSQNTRL---QPTP----- 60  
Db 662 SVSPQTSSTSESTSEPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPT 721  
QY 61 -----IPAPKPAAPRPLDRSPGVENKLIPIVSGSPASSTPLPDGTCGNSPTNNRA 115  
Db 722 EPTSEPTQPPAPQPPAPQPPAPQPPAPQPPAPQPPAPQPPAPQPPAPQPPAPQPPAPQPP 779  
QY 116 -----VTPVSGSNGSADPKAPPPVSSGSEPT-----PTLGNPDLGSLQ 155  
Db 780 PPKPTQPTSEPAQPTSESTSEPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQPTQ 834

RESULT 7  
Q57580 PRELIMINARY; PRT; 1151 AA.  
AC Q57580;  
DT 01-JUN-1998 (TREMELrel. 06, Created)  
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE High molecular mass nuclear antigen (Fragment).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98034404; PubMed=9365273;  
RA Shimada K.; Harata M.; Mizuno S.;  
RT "A nuclear matrix-associated high molecular mass nuclear antigen,

RT HMAA, of chicken and marked decrease of its immunoreactivity during

RL the progression of S phase.";  
 DR J. Cell Sci. 110:3031-3041(1997).  
 DR EMBL; D88440; BAA24137.1;  
 DR InterPro; IPR002965; P\_Fich\_extensn.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 FT NON\_TER 1 1151  
 FT NON\_TER 1 1151  
 SQ SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;

Query Match 15.8%; Score 162.5; DB 13; Length 1151;  
 Best Local Similarity 33.1%; Pred. No. 0.00052;  
 Matches 45; Conservative 12; Mismatches 40; Indels 39; Gaps 6;

QY 33 PKPLPQQPPANQDSSQNTLQTP-PIAPAPKPAAPRP-----LDRESQVE 84  
 DB 3 PTPPRNPTP-----PPAPSPAPAPATAPRPKWVFIAPLHPAPQPP 49  
 QY 85 NKLIPIVSGSPASSTPLPDGTGPNSTNNRAVTPVSGSNSSADPKAP---PPPPVSSG 141  
 DB 50 PKWVFIGAP-----PPPGTEP-----TPPSKPTDGADAAPKASAEITSPPPASP 94  
 QY 142 EPPTLGENPDGLSQSQ 157  
 DB 95 SPPDGFKAPSGAGEAE 110

## RESULT 8

QY6G51 PRELIMINARY; PRT; 816 AA.  
 AC Q96G51; Q96G51;  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Mitogen-activated protein kinase 7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas, and Muscle;  
 RA Strausberg R.;  
 RP Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; BC009963; AAH09963.1; -;  
 DR EMBL; BC007992; AAH07992.1; -;  
 DR EMBL; BC007404; AAH07404.1; -;  
 DR EMBL; BC030134; AAH030134.1; -;  
 DR HSSP; P24941; 1BUH.  
 DR InterPro; IPR003527; MAP\_kin.  
 DR InterPro; IPR00719; Prot\_kinase.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS01351; MAPK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 816 AA; 88386 MW; 27729FE31658CE3B CRC64;

Query Match 15.7%; Score 161.5; DB 4; Length 816;  
 Best Local Similarity 32.1%; Pred. No. 0.00044;  
 Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;

QY 17 RSTAPLNTQISALRNDPKLP-----QOPPAPANQDSSQNTLQTPPIP 63  
 DB 577 RPAAPALTSPAPAPATPTTPVQPTSPPPGPPVQAQGTGPPQAGSTSGVPQACPPP 636  
 QY 64 APAPKPAAPRPPLDRESFG-----VENKLIPIVSGSPASSTP-----LPPDGT 105  
 DB 637 GPAPHTGPPGPIPVAPPAQIATSTLSAAQSLVPPFGLFGSSTGVLVYFPFGLPPPD 696  
 QY 106 G--PNS-----TPNNRAVTPVSGSNSSADPKAPPPPPVSSGEPPTIG 147  
 DB 697 GGAPQSSMSSEPDVNLVT--QQLSKSQVEDFL-----PPVFSGTGPKGSG 738

## RESULT 9

QY6GT6 PRELIMINARY; PRT; 2321 AA.  
 AC Q9DGT6; Q9DGT6;  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
 DE RS1 immediate-early gene transactivator ICP4-like protein (RS1 immediate-early gene transactivator-like protein).  
 GN MDV084 OR MDV100.  
 OS Turkey herpesvirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Marek's disease-like viruses.  
 NCBI\_TaxID=10390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Md5;  
 RA MEDLINE=20392152; PubMed=10933706;  
 RX Tulman E.R., Afonso C.L., Lu Z., Zaak L., Rock D.L., Kutish G.F.;  
 RT "The genome of a very virulent Marek's disease virus.";  
 RL J. Virol. 74:7980-7988(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Md5;  
 RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Rock D.L., Kutish G.F.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF243438; AAG14284.1; -;  
 DR EMBL; AF243438; AAG14273.1; -;  
 DR InterPro; IPR005205; Herpes\_ICP4\_C.  
 DR InterPro; IPR005206; Herpes\_ICP4\_N.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR Pfam; PF03585; Herpes\_ICP4\_C; 1.  
 DR Pfam; PF03584; Herpes\_ICP4\_N; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 SQ SEQUENCE 2321 AA; 252959 MW; EF130BB0C4503003 CRC64;

Query Match 15.6%; Score 160.5; DB 12; Length 2321;  
 Best Local Similarity 32.4%; Pred. No. 0.0015;  
 Matches 60; Conservative 21; Mismatches 53; Indels 51; Gaps 13;

QY 12 NUKTERSTAPLNTQISALRNDPKLP-----POOPP-APANQD-----QNSQNTL- 56  
 DB 241 NRASERQLA--DTAASALR-APSPVFWAFDSRYPHLAPANQNSDPLCPETSTASAIL 297  
 QY 57 ---QPTPI---PAPAPKPAAP---PRPLDRSGVKNLIPIVSGSPAS-----ST 98  
 DB 298 HTNSPTPTSTSPAPISPTQPPACLPSPAPISPVQPPALLIPISTPTVFIQPS 357  
 QY 99 FLPPDGTGP-----NSTNNRAVTPVSGSNSS-----SADPKAPP---PPPVSSGPPPTL 146  
 DB 358 PSPPQAPSPPAHSSSCSPSHLAPSLSSPLSSPQLSPAPVSPSSPPPLSPGE--L 414  
 QY 147 GENPD 151  
 DB 415 APSPD 419

## RESULT 10

QY6G51 PRELIMINARY; PRT; 488 AA.  
 ID Q8DG51

Matches	52;	Conservative	13;	Mismatches	79;	Indels	22;	Gaps	7.
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

  

QY	33	KPK-LPOQPAPANQNSSNQTRLQP----	TPIIPAPA--PKPAAP-PRPLDRESGVE	84
DB	222	PTEPNAQTETAPATQAFQAQTETAPATAQTATVATPAPAAQAQRPAGAGSPAPAPAQANAPA	281	
QY	85	NKLIPSVCSPASSTLPPDDGTGNSTPNNRATVPVSQGS-NSSSADKPAPEPPPVSSEGEP	143	
DB	282	GSVVPEATVESITPAAPSAQTPTTRETACTEASPAPNSSAAPPNEPASEPV-AGRP	340	
QY	144	PTLGENDGLSQBQLSHRRERSLTQRDIQMFLPDKEFTGAQSGG	189	
DB	341	GTAASSSESASPVTTPRGET-----PDTAASAGTPSAG	374	

RESULT 12  
Q39492

ID	Q39492	PRELIMINARY;	PRT;	351 AA.
AC	Q39492;			
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)		
DE	WP6 protein precursor.			
OS	Chlamydomonas eugametos.			
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;			
OC	Chlamydomonadaceae; Chlamydomonas.			
NCBI_taxID=3053;				
[1]	SEQUENCE FROM N.A.			
RN	STRAIN=UTEX 10;			
RP	MEDLINE=95093034; PubMed=8000007;			
RX	Woessner J.P., Molendijk A.J., van Egmond P., Klis F.M.,			
RA	Goodenough U.W., Haring M.A.;			
RT	"domain conservation in several volvocalean cell wall proteins.";			
RL	Plant Mol. Biol. 26:947-960(1994).			
DR	EMBL; L29028; ABES3954.1; -.			
DR	InterPro; IPR002965; P rich extensin.			
DR	PRINTS; PR01217; PRICEXTENSIN.			
FT SIGNAL.				
FT CHAIN	1 31 POTENTIAL.			
FT FT	32 351 POTENTIAL.			
SQ SEQUENCE	351 AA; 70AD2EFFC74BB68 CRC64;			

Query Match 15.4%; Score 158.5; DB 10; Length 351;  
Best Local Similarity 30.9%; Pred. No. 0.00032;  
Matches 46; Conservative 15; Mismatches 71; Indels 17; Gaps 5;

QY	9	NISNNKTERTAPTINTQ----ISALRNDPKLPQQFPAPANAQDNS-----SQNTLRQPT	59
DB	144	SVYNXCNDTRPAPYNGCNSTFNVTTSVTTTPSPSPSPSPSPSPSPSPSPSKASPSKASPS	203
QY	60	P-PAPAPAKPAAAPRPLDRESPGVENKLIPSVGSPASTPLPPDGTFNSTPNNRATVP	118
DB	204	FSPKASPSKASPASPPSPQSPPTPSPKASPVASQQSPPTSPRSPTSPSTP---SP	258
QY	119	VSCQSNSSDADPKAPP--PPVSSGGPPT	145
DB	259	SPKASPPSPSPASPSLSLSPKYSPTSPT	287

RESULT 13  
Q9VZC2

ID	Q9VZC2	PRELIMINARY;	PRT;	420 AA.
AC	Q9VZC2; QBSZ47;			
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-OCT-2002	(TrEMBLrel. 22, Last sequence update)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)		
DE	Cg15021 protein (REL17165p).			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			

OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=2019606; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA George R.A., Lewis S.E., Berman B., Carlson J.W., Celniker S.E.,  
 RA Sutcliffe G.G., Workman J.R., Randell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Borkhan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Chertkov K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,  
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mervulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banson J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Murphy L., Nelson C.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]

RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003481; AA47902.2; -  
 DR EMBL; AY071124; AAL48746.1; -  
 DR FlyBase; FBGN0035544; CG15021.  
 DR InterPro; IPR003882; Pistil\_extensin.  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 DR PRINTS; PR01218; PSTEXTENSIN.  
 SQ SEQUENCE 420 AA; 42947 MW; 96D62FFDC9F996E1 CRC64;  
 Query Match 15.4%; Score 158.5; DB 5; Length 420;  
 Best Local Similarity 29.6%; Pred. No. 0.00038;  
 Matches 42; Conservative 11; Mismatches 64; Indels 25; Gaps 3;  
 QY 33 PKPLQPPAPANQDNSSQNTLRQTPPPAPAPA--PKPAAPPRDLRESGPGVENKLIIPS 90  
 DB 71 PQTTPPPPPPP-----QPTTPAPRSYGGPQTQPPRPPTPSAPAPPPPS 118  
 QY 91 VG-----SPASSTPLPPDGTGNTNNRAVTVSGGSSSSADPKAPPPPPVS 139  
 DB 119 YGPPQTPPPPPPPQTPSAFAPPSYGGPQTTPPPPPPTPSAPPSYGGPQPQPPAPQ 178  
 QY 140 SGEPTLGENPDGLSQSLQLEHR 161  
 DB 179 PPSPPSPQPGPEYLPDPQPKPR 200  
 RESULT 14  
 Q41805 PRELIMINARY; PRT; 1188 AA.  
 AC Q41805;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Extensin-like protein precursor.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 OC NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B73; TISSUE=Pollen;  
 RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;  
 RT "Pex genes: pollen-specific genes with extensin-like domains."  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z34465; CAA84230.1; -  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR007090; LRR\_plant.  
 DR Pfam; PF00560; LRR; 3.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 DR PROSITE; PS05052; LRR\_PS; 2.  
 KW Signal.  
 FT SIGNAL. 1 27 POTENTIAL.  
 SQ SEQUENCE 1188 AA; 120981 MW; 2C77C7F8D7130149 CRC64;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 14.2996 Seconds  
(without alignments)  
388.502 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQEQLHRSLSQTLRDIQRMFLPDEKEFTQA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq19Jun03.\*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	178	100.0	1426	23	Human legless homo
2	178	100.0	1435	22	Human BCL9 homolog
3	109	61.2	320	23	Mouse beta-catenin
4	109	61.2	1494	23	Mouse beta-catenin
5	107	60.1	738	23	Human beta-catenin
6	107	60.1	1115	23	Human legless homo
7	66	37.1	1429	22	Drosophila melanog
8	66	37.1	1464	23	D. melanogaster lg
9	57.5	32.3	603	24	Aspergillus fumiga

10	57.5	32.3	618	24	ABJ26453	Aspergillus fumiga
11	56	31.5	1294	22	ABR63502	Drosophila melanog
12	55	30.9	294	22	ABR95073	Human protein sequ
13	55	30.9	390	23	ABR96385	Human ovarian canc
14	55	30.9	433	22	ABR63951	Amino acid sequenc
15	55	30.9	433	22	ABR63952	Amino acid sequenc
16	55	30.9	433	22	ABR68522	Human GTP-binding
17	55	30.9	464	22	ABR92828	Human protein sequ
18	55	30.9	464	23	ABR96381	Human ovarian canc
19	55	30.9	643	23	ABR96382	Human ovarian canc
20	55	30.9	718	23	ABR96384	Human ovarian canc
21	55	30.9	751	23	ABR96383	Human ovarian canc
22	54.5	30.6	174	22	ABR6348	Murine bHLH transc
23	54.5	30.6	174	22	ABR6348	Murine bHLH transc
24	53	29.8	819	22	AAU69744	Thermus thermophil
25	52.5	29.5	181	22	ABR6345	Human bHLH transcr
26	52.5	29.5	181	22	ABR6347	Human bHLH transcr
27	52	29.2	959	21	AAU33051	Human secreted pro
28	51.5	28.9	210	23	ABP43969	Adenomatous polyo
29	51	28.7	329	21	AAU31371	Arabidopsis thalia
30	51	28.7	434	22	AAU33491	Enterococcus faeca
31	51	28.7	440	23	ABP47770	Protein #16 relate
32	51	28.7	448	22	AAU35058	Enterococcus faeca
33	51	28.7	500	22	ABR9346	Human secreted pro
34	51	28.7	1464	22	ABR71111	Drosophila melanog
35	51	28.7	1755	20	AAU41139	Mouse mammary tumo
36	50.5	28.4	96	22	ABR69662	Drosophila melanog
37	50.5	28.4	479	23	ABP43965	Unidentified prote
38	50.5	28.4	675	21	AAU54052	An angio genesis-as
39	50.5	28.4	675	21	AAU54053	A variant of an an
40	50.5	28.4	1183	22	ABR58769	Drosophila melanog
41	50	28.1	187	21	ABR38555	Human secreted pro
42	50	28.1	221	21	ABR07850	Amino acid sequenc
43	50	28.1	223	22	ABR62325	Glycine max glutat
44	50	28.1	275	21	ABR24369	Bacillus subtilis
45	50	28.1	342	20	AAU42781	Human neuronal imm

#### ALIGNMENTS

RESULT 1  
AAB71229  
ID AAB71229 standard; Protein; 1426 AA.  
XX  
AC AAB71229;  
XX  
DT 18-NOV-2002 (first entry)  
XX  
DE Human legless homologue lgs/bcl9 protein.

XX Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
XX tissue proliferation; tumour; cytosolic; cellular disorder; colon;  
XX blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
XX medulloblastoma; skin cancer; tissue regeneration; tissue repair.  
XX Homo sapiens.  
XX US2002086986-A1.  
XX  
XX 04-JUL-2002.  
XX  
XX 27-JUL-2001; 2001US-0915543.  
XX  
XX 28-JUL-2000; 2000US-221502P.  
XX

XX (BASL/) BASLER K.  
XX (BRUN/) BRUNNER E.  
XX (FROE/) FROESCH B.  
XX (KRAM/) KRAMPS T.  
XX (PETE/) PETER O.

XX Basler K, Brunner E, Froesch B, Kramps T, Peter O;

XX WPI: 2002-635689/68.  
 DR N-PSDB; AAF88467.  
 XX  
 PT Novel polypeptide useful in therapeutic method for treating disorders  
 PT of cell fate such as cell differentiation or cell proliferation -  
 XX  
 PS Example II; Fig 8B; 41pp; English.  
 XX  
 CC This invention describes a novel polypeptide sharing one or more  
 CC homologous amino acid domains with the legless (lgs) protein, a  
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
 CC involved in the formation and maintenance of spatial arrangements  
 CC and proliferation of tissues during development, and in the formation  
 CC and growth of many human tumours. The products of the invention have  
 CC cytostatic activity and can be used to treat cellular disorders, blood  
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,  
 CC where the cancerous condition is colon, breast, head and neck, brain,  
 CC thyroid, medulloblastoma or skin cancer. The product could also be used  
 CC to promote tissue regeneration and repair. This sequence represents the  
 CC human legless (lgs) protein homologue lgs/bcl9 described in the  
 CC disclosure of the invention.  
 XX  
 SQ Sequence 1426 AA;  
 Query Match 100.0%; Score 178; DB 23; Length 1426;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-16;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGLSQEQLHRSLSQTLRDIQRLMFPDEKEFTGA 35  
 Db 349 DGLSQEQLHRSLSQTLRDIQRLMFPDEKEFTGA 383  
 RESULT 2  
 ID ABB11808 standard; peptide; 1435 AA.  
 AC ABB11808;  
 DT 11-JAN-2002 (first entry)  
 XX  
 DE Human BCL9 homologue, SEQ ID NO:2178.  
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antiulcer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157188-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX

DR WPI: 2001-457740/49.  
 DR N-PSDB; ABA09052.  
 XX  
 PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 PS Claim 20; Page 256-257; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 1435 AA;  
 Query Match 100.0%; Score 178; DB 22; Length 1435;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-16;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGLSQEQLHRSLSQTLRDIQRLMFPDEKEFTGA 35  
 Db 389 DGLSQEQLHRSLSQTLRDIQRLMFPDEKEFTGA 423  
 RESULT 3  
 AAU78461  
 ID AAU78461 standard; Protein; 320 AA.  
 XX  
 AC AAU78461;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Mouse beta-catenin nuclear localised protein #2.  
 XX  
 DE Mouse; beta-catenin nuclear localised protein; cancer;  
 KW Gene therapy; EST; expressed sequence tag.  
 XX  
 OS Mus musculus.  
 XX



```
PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP08140.
XX
PR 22-SEP-2000; 2000JP-0287876.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
PWPI; 2002-330014/36.
DR N-PSDB; ABK47632.
XX
New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer.
XX
PS Claim 2; Page 91-92; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC mouse beta-catenin nuclear localised protein #2.
XX
SQ Sequence 320 AA;
XX
Query Match 61.2%; Score 109; DB 23; Length 320;
Best Local Similarity 84.0%; Pred. No. 6.3e-07;
Matches 21; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 DGLSQQLHRSLSQTLRDQRL 25
DB 150 EGLSQQLHRSLSQTLRDIERLL 174
XX
RESULT 4
AAU78460
ID AAU78460 standard; Protein; 1494 AA.
XX
AC AAU78460;
XX
DT 02-JUL-2002 (first entry)
XX
DE Mouse beta-catenin nuclear localised protein.
XX
KW Mouse; beta-catenin nuclear localised protein; cancer;
XX gene therapy; EST; expressed sequence tag.
XX
OS Mus musculus.
XX
PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP08140.
XX
PR 22-SEP-2000; 2000JP-0287876.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
PWPI; 2002-330014/36.
DR N-PSDB; ABK47631.
XX
New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer.
XX
PS Claim 2; Page 91-92; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC mouse beta-catenin nuclear localised protein #2.
XX
SQ Sequence 320 AA;
XX
Query Match 61.2%; Score 109; DB 23; Length 320;
Best Local Similarity 84.0%; Pred. No. 6.3e-07;
Matches 21; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 DGLSQQLHRSLSQTLRDQRL 25
DB 150 EGLSQQLHRSLSQTLRDIERLL 174
XX
RESULT 5
AAU78463
ID AAU78463 standard; Protein; 738 AA.
XX
AC AAU78463;
XX
DT 02-JUL-2002 (first entry)
XX
DE Human beta-catenin nuclear localised protein #2.
XX
KW Mouse; beta-catenin nuclear localised protein; cancer;
XX gene therapy; EST; expressed sequence tag.
XX
OS Homo sapiens.
XX
PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP08140.
XX
PR 22-SEP-2000; 2000JP-0287876.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
PWPI; 2002-330014/36.
DR N-PSDB; ABK47638.
XX
New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer.
XX
PS Claim 8; Page 102-105; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC human beta-catenin nuclear localised protein #2.
XX
SQ Sequence 738 AA;
XX
Query Match 60.1%; Score 107; DB 23; Length 738;
Best Local Similarity 87.5%; Pred. No. 3e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
XX
QY 2 GLSQQLHRSLSQTLRDQRL 25
DB 1 GLSQQLHRSLSQTLRDIERLL 24
XX
```

RESULT 6  
ID AAB71230 standard; Protein; 1115 AA.  
XX AC AAB71230;  
XX DT 18-NOV-2002 (first entry)  
XX DE Human legless homologue higs-1 partial protein.  
XX KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
XX KW tissue proliferation; tumour; cytoskeletal; cellular disorder; colon;  
XX KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
XX KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.  
XX OS Homo sapiens.  
XX FN US2002086986-A1.  
XX PD 04-JUL-2002.  
XX PF 27-JUL-2001; 2001US-0915543.  
XX PR 28-JUL-2000; 2000US-221502P.  
XX PA (BASL/) BASLER K.  
XX PA (BRUN/) BRUNNER E.  
XX PA (FROE/) FROESCH B.  
XX PA (KRAM/) KRAMPS T.  
XX PA (PETE/) PETER O.  
XX PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;  
XX DR WPI: 2002-635689/68.  
XX DR N-PSDB; AAF88468.  
XX PT Novel polypeptide useful in therapeutic method for treating disorders  
XX PT of cell fate such as cell differentiation or cell proliferation -  
XX PS Example II; Fig 10B; 41pp; English.  
XX CC This invention describes a novel polypeptide sharing one or more  
XX CC homologous amino acid domains with the legless (lgs) protein, a  
XX CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
XX CC involved in the formation and maintenance of spatial arrangements  
XX CC and proliferation of tissues during development, and in the formation  
XX CC and growth of many human tumours. The products of the invention have  
XX CC cytoskeletal activity and can be used to treat cellular disorders, blood  
XX CC disorders and cancers caused by over-stimulation of the Wnt pathway,  
XX CC where the cancerous condition is colon, breast, head and neck, brain,  
XX CC thyroid, medulloblastoma or skin cancer. The product could also be used  
XX CC to promote tissue regeneration and repair. This sequence represents the  
XX CC human legless (lgs) protein homologue higs-1 described in the  
XX CC disclosure of the invention.  
SQ Sequence 1115 AA;  
Query Match 60.1%; Score 107; DB 23; Length 1115;  
Best Local Similarity 87.5%; Pred. No. 4.8e-06;  
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 GLSQEQLEHRSRLQTLRDQRL 25  
Db 76 GLSQEQLEHRSRLQTLRDQRL 99  
RESULT 7  
ID ABB58779 standard; Protein; 1429 AA.  
XX AC ABB58779;  
XX PN ABB58779;  
XX XX

DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 3129.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX FN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX DR WPI: 2001-656860/75.  
XX DR N-PSDB; ABL02882.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX PT interactions -  
XX PS Disclosure; SEQ ID NO 3129; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX CC sequences (ABL01840-ABL16175) and the encoded proteins  
XX CC (ABBS7737-ABBS72072).  
XX CC The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 1429 AA;  
Query Match 37.1%; Score 66; DB 22; Length 1429;  
Best Local Similarity 31.4%; Pred. No. 3.1;  
Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
Qy 1 DGLSQEQLEHRSRLQTLRDQRLPPDEKFTGA 35  
Db 520 ENLTPQQRHREQLAKIKKMQFLPPENSVGA 554  
RESULT 8  
ID AAB71228 standard; Protein; 1464 AA.  
XX AC AAB71228;  
XX DT 18-NOV-2002 (first entry)  
XX DE D. melanogaster lgs protein.  
XX KW Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
XX KW tissue proliferation; tumour; cytoskeletal; cellular disorder; colon;  
XX KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
XX KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.  
XX OS Drosophila melanogaster.  
XX FN US2002086986-A1.  
XX XX

PD 04-JUL-2002.  
 XX  
 PF 27-JUL-2001; 2001US-0915543.  
 XX  
 PR 28-JUL-2000; 2000US-221502P.  
 XX  
 PA (BASL/) BASLER K.  
 PA (BRUN/) BRUNNER E.  
 PA (FROE/) FROESCH B.  
 PA (KRAM/) KRAMPS T.  
 PA (PETE/) PETER O.  
 XX  
 PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;  
 XX  
 DR WPI; 2002-635689/68.  
 DR N-PSDB; AAF88466.  
 XX  
 PT Novel polypeptide useful in therapeutic method for treating disorders  
 PT of cell fate such as cell differentiation or cell proliferation -  
 XX  
 PS Example II; Fig 2; 41pp; English.  
 XX  
 CC This invention describes a novel polypeptide sharing one or more  
 CC homologous amino acid domains with the legless (lgs) protein, a  
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
 CC involved in the formation and maintenance of spatial arrangements  
 CC and proliferation of tissues during development, and in the formation  
 CC and growth of many human tumors. The products of the invention have  
 CC cytotostatic activity and can be used to treat cellular disorders, blood  
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,  
 CC where the cancerous condition is colon, breast, head and neck, brain,  
 CC thyroid, medulloblastoma or skin cancer. The product could also be used  
 CC to promote tissue regeneration and repair. This sequence represents the  
 CC Drosophila melanogaster (fruitfly) legless (lgs) protein described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 1464 AA;  
 Query Match 37.1%; Score 66; DB 23; Length 1464;  
 Best Local Similarity 31.4%; Pred. No. 3.2;  
 Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 DGLSQEQLEHRRSLQTLRDIOQMFLPDEKEFTGA 35  
 DB 515 ENLTQQRHREQLAKIKKQNFLEPENSVA 549  
 RESULT 9  
 ABJ25853  
 ID ABJ25853 standard; Protein; 603 AA.  
 XX  
 AC ABJ25853;  
 XX  
 DT 16-APR-2003 (first entry)  
 XX  
 DE Aspergillus fumigatus essential gene protein #511.  
 XX  
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
 KW cancer; contamination; biofilm; antibody; immune response.  
 XX  
 OS Aspergillus fumigatus.  
 XX  
 PN WO200286090-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 23-APR-2002; 2002WO-US13142.  
 XX  
 PR 23-APR-2001; 2001US-285697P.  
 PR 27-APR-2001; 2001US-287066P.  
 PR 05-JUN-2001; 2001US-295890P.  
 PR 09-JUL-2001; 2001US-303899P.  
 PR 31-AUG-2001; 2001US-316362P.  
 XX

XX PA (BLIT-) ELITRA PHARM INC.  
 XX  
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
 XX  
 DR WPI; 2003-093124/08.  
 XX  
 PT New purified or isolated nucleic acids of essential genes of  
 PT Aspergillus fumigatus, useful for treating or preventing infections by  
 PT A. fumigatus, or for treating a non-infectious disease in a subject  
 PT e.g. cancer -  
 XX  
 PS Disclosure; Page -; 175pp; English.  
 XX  
 CC The invention relates to novel purified or isolated nucleic acids of  
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of  
 CC the invention are used to treat or prevent infections by a pathogenic  
 CC organism such as A. fumigatus, to treat a non-infectious disease in a  
 CC subject (e.g. cancer), to prevent or contain contamination of an object  
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a  
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for  
 CC expressing recombinant protein for characterisation, screening or  
 CC therapeutic use, as markers for host tissues in which the pathogenic  
 CC organisms invade or reside, for comparing with the DNA sequence of A.  
 CC fumigatus to identify duplicated genes or paralogues having the same or  
 CC similar biochemical activity and/or function, for comparing with DNA  
 CC sequences of other related or distant pathogenic organisms to identify  
 CC potential orthologous essential or virulence genes, for selecting and  
 CC making oligomers for attachment to a nucleic acid array for examination  
 CC of expression patterns, for raising anti-protein antibodies, as an  
 CC antigen to raise anti-DNA antibodies or to elicit another immune  
 CC response, and for identifying polynucleotides encoding the other protein  
 CC with which binding occurs or to identify inhibitors of the binding  
 CC interaction. The polypeptides may be used to raise antibodies or to  
 CC elicit immune response, as a reagent in assays designed to quantitatively  
 CC determine levels of the protein in biological fluids, as a marker for  
 CC host tissues in which pathogenic organism invade or reside, and to  
 CC isolate correlative receptors or ligands in the case of virulence  
 CC factors. This sequence represents a protein of one of the essential genes  
 CC of Aspergillus fumigatus of the invention.  
 XX  
 SQ Sequence 603 AA;  
 Query Match 32.3%; Score 57.5; DB 24; Length 603;  
 Best Local Similarity 31.4%; Pred. No. 18;  
 Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;  
 QY 1 DGLSQEQLEHRRSLQTLRDIOQMFLPDEKEFTGA 35  
 DB 45 DGVETEKIREKD---EVEKKLERMLFGDDEGFGA 76  
 RESULT 10  
 ABJ26453  
 ID ABJ26453 standard; Protein; 618 AA.  
 XX  
 AC ABJ26453;  
 XX  
 DT 16-APR-2003 (first entry)  
 XX  
 DE Aspergillus fumigatus essential gene protein #111.  
 XX  
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
 KW cancer; contamination; biofilm; antibody; immune response.  
 XX  
 OS Aspergillus fumigatus.  
 XX  
 PN WO200286090-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 23-APR-2002; 2002WO-US13142.  
 XX



PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 FT full-length cDNAs -  
 FT  
 FT Claim 8; SEQ ID 16943; 2537pp + CD ROM; English.  
 PS  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 294 AA;  
 Query Match 30.9%; Score 55; DB 22; Length 294;  
 Best Local Similarity 52.2%; Pred. No. 18;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQQLHRRSRLQTLRDQRM 24  
 DB 224 GLRTEGLFRRSASVQTVREIQRL 246  
 RESULT 13  
 ID ABG96285 standard; Protein; 390 AA.  
 XX  
 AC ABG96285;  
 XX  
 XX 11-DEC-2002 (first entry)  
 DT  
 XX Human ovarian cancer marker M355.  
 DE  
 XX Human; ovarian cancer; marker; cancer; familial history; brain disorder;  
 KW central nervous system disorder; bacterial meningitis; viral meningitis;  
 KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;  
 KW brain herniation; inflammation; encephalitis; testicular disorder;  
 KW nontuberculous granulomatous orchitis; connective tissue disorder;  
 KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;  
 KW histological type; carcinogenic; ovarian cancer marker.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200271928-A2.  
 PN  
 XX  
 PD 19-SEP-2002.  
 XX  
 XX 14-MAR-2002; 2002WO-US07825.  
 PF  
 XX 14-MAR-2001; 2001US-276025P.  
 PR 14-MAR-2001; 2001US-276026P.  
 PR

10-AUG-2001; 2001US-311732P.  
 19-SEP-2001; 2001US-323580P.  
 26-SEP-2001; 2001US-324967P.  
 26-SEP-2001; 2001US-325102P.  
 26-SEP-2001; 2001US-325149P.  
 (MILL-) MILLENNIUM PHARM INC.  
 Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;  
 Meyers RE, Morrissy MP, Olandt PJ, Sen A, Vieby PO, Mills GB;  
 Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;  
 WPI; 2002-723277/78.  
 N-PSDB; ABS76377.  
 Assessing whether a patient is afflicted with ovarian cancer, useful in  
 assessing the stage or progression of the disease, comprises comparing  
 the expression level of a cancer marker in a sample from a patient and  
 from a non cancer patient -  
 Disclosure; Page 134-135; 481pp; English.  
 The present invention relates to a new method for assessing whether a  
 patient is afflicted with ovarian cancer. The method involves comparing  
 the expression level of a marker in a patient sample and the normal level  
 of expression of the marker in a control non-ovarian cancer sample, where  
 the marker is selected from 363 cancer markers described in the  
 specification. The method of the invention is useful in diagnosing or  
 characterising cancer, in detecting the presence of cancer as early as  
 possible, and the recurrence of ovarian cancer. The method may also be of  
 particular use with patients having an enhanced risk of developing  
 ovarian cancer (e.g. patients having a familial history of ovarian  
 cancer). The cancer markers may be used in the management and treatment  
 of e.g. brain and central nervous system disorders (e.g. bacterial and  
 viral meningitis, Alzheimer's disease or Parkinson's disease), brain  
 disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),  
 inflammations (e.g. bacterial or viral meningitis or encephalitis),  
 testicular disorders (e.g. nontuberculous granulomatous orchitis),  
 connective tissue disorders, or heart disorders (e.g. ischaemic heart  
 disease or atherosclerosis). The compositions and methods may also be  
 used in assessing the histological type of neoplasm associated with  
 ovarian cancer, monitoring the progression of ovarian cancer,  
 determining whether ovarian cancer has metastasised or is likely to  
 metastasize, selecting a composition for inhibiting ovarian cancer,  
 assessing the ovarian carcinogenic potential of a compound, or  
 inhibiting ovarian cancer or at risk of developing ovarian cancer. The  
 present amino acid sequence represents one of the ovarian cancer markers  
 described in the invention.  
 Sequence 390 AA;  
 Query Match 30.9%; Score 55; DB 23; Length 390;  
 Best Local Similarity 52.2%; Pred. No. 25;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQQLHRRSRLQTLRDQRM 24  
 DB 181 GLRTEGLFRRSASVQTVREIQRL 203  
 RESULT 14  
 ID AAG63851 standard; Protein; 433 AA.  
 XX  
 AC AAG63851;  
 XX  
 XX 29-OCT-2001 (first entry)  
 DT  
 XX Amino acid sequence of human GTPase activating protein GTPAP1.  
 DE  
 XX GTPase activating protein; GTPAP1; cell signalling; immune disorder;  
 KW cell proliferative disorder; cancer; colon cancer; arteriosclerosis;  
 KW diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 60 /note= "potential protein kinase phosphorylation site for tyrosine kinase"

FT Modified-site 101 /note= "potential protein kinase phosphorylation site for tyrosine kinase"

FT Modified-site 129 /note= "potential protein kinase phosphorylation site for protein kinase C"

FT Modified-site 169 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Domain 210..235 /note= "GTPase-activator domain"

FT Domain 310..350 /note= "GTPase-activator domain"

FT Modified-site 239 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Modified-site 239 /note= "potential protein kinase phosphorylation site for protein kinase C"

FT Modified-site 292 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Modified-site 297 /note= "potential protein kinase phosphorylation site for protein kinase C"

FT Modified-site 309 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Domain 310..350 /note= "GTPase-activator domain"

FT Modified-site 315 /note= "potential protein kinase phosphorylation site for tyrosine kinase"

FT Modified-site 338 /note= "potential N-linked glycosylation site"

FT Modified-site 382 /note= "potential protein kinase phosphorylation site for casein kinase II"

XX WO200161010-A2.

XX 23-AUG-2001.

XX 15-FEB-2001; 2001WO-US05075.

XX 18-FEB-2000; 2000US-0507765.

XX (INCY-) INCYTE GENOMICS INC.

XX Klingler TM, Stewart EA, Yue H, Baughn MR;

XX WPI: 2001-522598/57.

XX N-PSDB: AAH74978.

XX A substantially purified GTPase activating protein useful for treating or preventing cell signalling, immune and cell proliferative disorders, including cancer especially colon cancer -

XX Claim 14; Fig 1A-E; 78pp; English.

XX The present sequence represents a human GTPase activating protein, designated GTPAP1. GTPAP polynucleotides and polypeptides are useful for treating or preventing a disease or condition associated with altered expression of GTPase activating proteins, especially cell signalling, immune and cell proliferative disorders, including cancer especially colon cancer. Examples of diseases treatable include

CC arteriosclerosis, diabetes, psoriasis, hepatitis and multiple sclerosis.

CC GTPAP1 polynucleotides are also useful for gene therapy treatments of the diseases.

XX SQ Sequence 433 AA;

Query Match 30.9%; Score 55; DB 22; Length 433;

Best Local Similarity 52.2%; Pred. No. 28;

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQEQLHRRSLQTLRDIRM 24

DB 224 GLRTEGLFRSASVQTVREIQRL 246

RESULT 15

AAAG63852

ID AAG63852 standard; Protein; 433 AA.

XX AC AAG63852;

XX 29-OCT-2001 (first entry)

XX Amino acid sequence of human GTPase activating protein GTPAP2.

DE GTPase activating protein; GTPAP2; cell signalling; immune disorder;

KW cell proliferative disorder; cancer; colon cancer; arteriosclerosis;

KW diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 60 /note= "potential protein kinase phosphorylation site for tyrosine kinase"

FT Modified-site 101 /note= "potential protein kinase phosphorylation site for tyrosine kinase"

FT Modified-site 129 /note= "potential protein kinase phosphorylation site for protein kinase C"

FT Modified-site 169 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Domain 210..235 /note= "GTPase-activator domain"

FT Domain 310..350 /note= "GTPase-activator domain"

FT Modified-site 239 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Modified-site 239 /note= "potential protein kinase phosphorylation site for protein kinase C"

FT Modified-site 292 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Modified-site 297 /note= "potential protein kinase phosphorylation site for protein kinase C"

FT Modified-site 309 /note= "potential protein kinase phosphorylation site for casein kinase II"

FT Domain 310..350 /note= "GTPase-activator domain"

FT Modified-site 315 /note= "potential protein kinase phosphorylation site for tyrosine kinase"

FT Modified-site 338 /note= "potential N-linked glycosylation site"

FT Modified-site 382 /note= "potential protein kinase phosphorylation site for casein kinase II"

XX WO200161010-A2.  
 XX  
 XX PD 23-AUG-2001.  
 XX  
 XX PF 15-FEB-2001; 2001WO-US05075.  
 XX  
 XX PR 18-FEB-2000; 2000US-0507765.  
 XX  
 XX PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX PI Klingler TM, Stewart EA, Yue H, Baughn MR;  
 XX  
 XX DR WPI; 2001-522598/57.  
 XX  
 XX DR N-PSDB; AAH74979.  
 XX  
 XX PT A substantially purified GTPase activating protein useful for treating  
 PT or preventing cell signalling, immune and cell proliferative disorders,  
 PT including cancer especially colon cancer -  
 XX  
 XX PS Claim 14; Fig 2A-E; 78pp; English.  
 XX  
 XX CC The present sequence represents a human GTPase activating protein,  
 CC designated GTPAP2. GTPAP polynucleotides and polypeptides are useful  
 CC for treating or preventing a disease or condition associated with  
 CC altered expression of GTPase activating proteins, especially cell  
 CC signalling, immune and cell proliferative disorders, including cancer  
 CC especially colon cancer. Examples of diseases treatable include  
 CC arteriosclerosis, diabetes, psoriasis, hepatitis and multiple sclerosis.  
 CC GTPAP polynucleotides are also useful for gene therapy treatments of  
 CC the diseases.  
 XX  
 XX SQ Sequence 433 AA;  
 Query Match 30.9%; Score 55; DB 22; Length 433;  
 Best Local Similarity 52.2%; Pred. No; 28;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQQLHRRSLQTLRDIQRM 24  
 DB 224 GLRTEGLFRSASVQTVREIQRL 246

Search completed: November 13, 2003, 09:25:20  
 Job time : 15.296 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:19 ; Search time 5.1751 Seconds  
(without alignments)  
286.155 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178  
Sequence: 1 DGLSQQLHRSLSQTLRDIQRLPFDEKEFTGA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	55	30.9	333	4	US-09-507-765-32
2	55	30.9	433	4	US-09-507-765-30
3	55	30.9	433	4	US-09-507-765-31
4	54.5	30.6	1242	4	US-09-252-991A-25843
5	53	29.8	819	4	US-09-651-656-15
6	53	29.8	819	4	US-09-650-855-15
7	52	29.2	578	4	US-09-252-991A-23906
8	51	28.7	589	4	US-09-252-991A-24834
9	50.5	28.4	354	4	US-09-252-991A-28779
10	50	28.1	221	3	US-09-247-373B-54
11	50	28.1	275	4	US-09-542-749A-2
12	50	28.1	1201	4	US-09-252-991A-32259
13	49	27.5	524	4	US-09-186-276B-56
14	49	27.5	524	4	US-08-842-445-56
15	49	27.5	524	4	US-09-186-188B-56
16	49	27.5	1027	4	US-09-252-991A-23210
17	48.5	27.2	1872	1	US-08-188-582-14
18	48.5	27.2	1872	1	US-08-646-715-11
19	48.5	27.2	1893	1	US-08-188-582-11
20	48.5	27.2	1893	1	US-08-646-715-11
21	48	27.0	266	4	US-09-107-532A-5226
22	48	27.0	1097	4	US-09-252-991A-22579
23	48	27.0	1315	4	US-09-252-991A-22746
24	47.5	26.7	846	4	US-09-252-991A-29246
25	47	26.4	1000	4	US-09-252-991A-31361
26	47	26.4	10182	4	US-09-134-001C-3159
27	46.5	26.1	65	1	US-08-227-536-3

Sequence 3, Appli  
Sequence 5104, Ap  
Sequence 3753, Ap  
Sequence 18906, A  
Sequence 31012, A  
Sequence 23336, A  
Sequence 6561, Ap  
Sequence 16908, A  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 385, App  
Sequence 19252, A  
Sequence 34, Appli

28 46.5 26.1 65 5 PCT-US95-04682-3  
29 46.5 26.1 885 4 US-09-107-532A-5104  
30 46.5 26.1 935 4 US-09-107-532A-3753  
31 46 25.8 178 4 US-09-252-991A-18906  
32 46 25.8 180 4 US-09-252-991A-31012  
33 46 25.8 452 4 US-09-252-991A-23336  
34 46 25.8 465 4 US-09-328-352-6561  
35 46 25.8 558 4 US-09-252-991A-16908  
36 46 25.8 2101 1 US-08-466-390-4  
37 46 25.8 2101 1 US-08-470-950-4  
38 46 25.8 2101 1 US-08-467-781-4  
39 46 25.8 2101 1 US-08-195-487-4  
40 46 25.8 2101 2 US-08-483-924-4  
41 46 25.8 2101 3 US-09-452-294-1  
42 46 25.8 2101 5 PCT-US93-06160-4  
43 45.5 25.6 118 4 US-08-858-207A-395  
44 45.5 25.6 832 4 US-09-252-991A-19252  
45 45 25.3 52 2 US-08-690-011A-34

## ALIGNMENTS

RESULT 1  
US-09-507-765-32  
; Sequence 32, Application US/09507765  
; Patent No. 6509155  
; GENERAL INFORMATION:  
; APPLICANT: Klingner, Tod M.  
; APPLICANT: Stewart, Elizabeth A.  
; APPLICANT: Yue, Henry  
; APPLICANT: Baughn, Mariah R.  
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
; FILE REFERENCE: PC-0010 US  
; CURRENT APPLICATION NUMBER: US/09/507,765  
; CURRENT FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PERL Program  
; SEQ ID NO 32  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. 6509155 g6572185  
US-09-507-765-32

Query Match 30.9%; Score 55; DB 4; Length 333;  
Best Local Similarity 52.2%; Pred No. 6.9;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSQTLRDIQRM 24  
Db 124 GLRTEGLFRSSASVQTVREIQRL 146

RESULT 2  
US-09-507-765-30  
; Sequence 30, Application US/09507765  
; Patent No. 6509155  
; GENERAL INFORMATION:  
; APPLICANT: Klingner, Tod M.  
; APPLICANT: Stewart, Elizabeth A.  
; APPLICANT: Yue, Henry  
; APPLICANT: Baughn, Mariah R.  
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
; FILE REFERENCE: PC-0010 US  
; CURRENT APPLICATION NUMBER: US/09/507,765  
; CURRENT FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PERL Program  
; SEQ ID NO 30  
; LENGTH: 433



TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6509155 3068538CD1  
US-09-507-765-30

Query Match 30.9%; Score 55; DB 4; Length 433;  
Best Local Similarity 52.2%; Pred. No. 9.1;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDICRM 24  
DB 224 GLRTEGLFRSGASQTVREIQRL 246

RESULT 3  
US-09-507-765-31  
; Sequence 31, Application US/09507765  
; Patent No. 6509155  
; GENERAL INFORMATION:  
; APPLICANT: Klingner, Tod M.  
; APPLICANT: Stewart, Elizabeth A.  
; APPLICANT: Yue, Henry  
; APPLICANT: Baughn, Maria R.  
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
; FILE REFERENCE: PC-0010 US  
; CURRENT APPLICATION NUMBER: US/09/507,765  
; CURRENT FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PERL Program  
; SEQ ID NO 31  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6509155 404424.5.pseq  
US-09-507-765-31

Query Match 30.9%; Score 55; DB 4; Length 433;  
Best Local Similarity 52.2%; Pred. No. 9.1;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDICRM 24  
DB 224 GLRTEGLFRSGASQTVREIQRL 246

RESULT 4  
US-09-252-991A-25843  
; Sequence 25843, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25843  
; LENGTH: 1242  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25843

Query Match 30.6%; Score 54.5; DB 4; Length 1242;  
Best Local Similarity 35.1%; Pred. No. 33;

Matches 13; Conservative 9; Mismatches 6; Indels 9; Gaps 1;  
QY 5 EQLEHRRSL-----QTLRDICRMILFPDKEP 32  
DB 169 RQLHRRQSLRSIDIAIEIRSLDLRLLLFIYQREF 205

RESULT 5  
US-09-651-656-15  
; Sequence 15, Application US/09651656  
; Patent No. 6340566  
; GENERAL INFORMATION:  
; APPLICANT: MCCUTHEN-MALONEY, SANDRA  
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE  
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,  
; TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES  
; FILE REFERENCE: IL-10689  
; CURRENT APPLICATION NUMBER: US/09/651,656  
; CURRENT FILING DATE: 2000-08-29  
; PRIOR APPLICATION NUMBER: 60/192,764  
; PRIOR FILING DATE: 2000-03-28  
; NUMBER OF SEQ ID NOS: 106  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 819  
; TYPE: PRT  
; ORGANISM: Thermus thermophilus  
US-09-651-656-15

Query Match 29.8%; Score 53; DB 4; Length 819;  
Best Local Similarity 42.3%; Pred. No. 34;  
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLEHRRSLQTLRDICRMILFPDKE 31  
DB 466 EKVPQYRPVQTLKDRQRYTLPEMKE 491

RESULT 6  
US-09-650-855-15  
; Sequence 15, Application US/09650855  
; Patent No. 6365355  
; GENERAL INFORMATION:  
; APPLICANT: MCCUTHEN-MALONEY, SANDRA  
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA  
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA  
; TITLE OF INVENTION: MISMATCHES  
; FILE REFERENCE: IL-10284  
; CURRENT APPLICATION NUMBER: US/09/650,855  
; CURRENT FILING DATE: 2000-08-29  
; PRIOR APPLICATION NUMBER: 60/192,764  
; PRIOR FILING DATE: 2000-03-28  
; NUMBER OF SEQ ID NOS: 106  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 819  
; TYPE: PRT  
; ORGANISM: Thermus thermophilus  
US-09-650-855-15

Query Match 29.8%; Score 53; DB 4; Length 819;  
Best Local Similarity 42.3%; Pred. No. 34;  
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLEHRRSLQTLRDICRMILFPDKE 31  
DB 466 EKVPQYRPVQTLKDRQRYTLPEMKE 491

RESULT 7  
US-09-252-991A-23906

```
; Sequence 23906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23906
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23906

Query Match      29.2%; Score 52; DB 4; Length 578;
Best Local Similarity 84.6%; Pred. No. 32;
Matches 11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      6 EQLEHRSLSQTL 18
Db      339 ECLHRSLSRTL 351

RESULT 8
US-09-252-991A-24834
; Sequence 24834, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24834
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24834

Query Match      28.7%; Score 51; DB 4; Length 589;
Best Local Similarity 37.0%; Pred. No. 44;
Matches 10; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY      5 EQLEHRSLSQTL 31
Db      156 EQQRHRSLSQTL 182

RESULT 9
US-09-252-991A-28779
; Sequence 28779, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

```
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28779
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28779

Query Match      28.4%; Score 50.5; DB 4; Length 354;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 14; Conservative 6; Mismatches 8; Indels 7; Gaps 2;

QY      2 GLSQE-----QLHRSLSQTL 31
Db      252 GLAQRLLPGAPHRPRTLAVDRQRR--FDLRQ 284

RESULT 10
US-09-247-373B-54
; Sequence 54, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 221
; TYPE: PRT
; ORGANISM: SOYBEAN
US-09-247-373B-54

Query Match      28.1%; Score 50; DB 3; Length 221;
Best Local Similarity 36.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 8; Indels 4; Gaps 1;

QY      9 EHRERSLQTL---LRDIQRLMFPDEKFTG 34
Db      116 EEREKSIKIWEHLRVVVENQCFGDKKPF 145

RESULT 11
US-09-542-749A-2
; Sequence 2, Application US/09542749A
; Patent No. 6428981
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Steen T.
; APPLICANT: Christensen, Christina L.
; APPLICANT: Kristensen, Tina
; TITLE OF INVENTION: A Bacillus Protein Production Cell
; FILE REFERENCE: 5861.200-US
; CURRENT APPLICATION NUMBER: US/09/542,749A
; CURRENT FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: 60/130,194
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: PA 1999 00506
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-542-749A-2

Query Match      28.1%; Score 50; DB 4; Length 275;
```



Thu Nov 13 10:41:03 2003

us-09-915-543-15\_copy\_349\_383.ra1

Page 5

Job time : 6.1751 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:27:44 ; Search time 9.39689 Seconds  
(without alignments)  
679.968 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383  
Perfect score: 178  
Sequence: 1 DGLSQEQLHRSLSQTLRDIQRLFPDEKEFTGA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues  
Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	35	15	US-10-322-579-5
2	178	100.0	1426	15	US-10-322-579-15
3	107	60.1	1115	15	US-10-322-579-17
4	66	37.1	35	15	US-10-322-579-4
5	57.5	32.3	603	15	US-10-128-714-3511
6	57.5	32.3	618	15	US-10-128-714-8511
7	55	30.9	294	12	US-10-053-248-12
8	55	30.9	333	16	US-10-284-753-32
9	55	30.9	390	15	US-10-097-340-21
10	55	30.9	433	16	US-10-284-753-30
11	55	30.9	433	16	US-10-284-753-31
12	55	30.9	464	15	US-10-097-340-14
13	55	30.9	643	15	US-10-097-340-16
14	55	30.9	718	15	US-10-097-340-19
15	55	30.9	751	15	US-10-097-340-18

16	51	28.7	434	9	US-09-815-242-4987	Sequence 1987, Ap
17	51	28.7	448	9	US-09-815-242-10651	Sequence 10651, A
18	51	28.7	500	10	US-09-731-872-466	Sequence 466, App
19	51	28.7	500	12	US-09-876-997-466	Sequence 466, App
20	50	28.1	342	12	US-09-244-805-62	Sequence 62, Appl
21	50	28.1	444	10	US-09-738-626-4366	Sequence 4366, Ap
22	50	28.1	1162	9	US-09-815-242-11828	Sequence 11828, A
23	49.5	27.8	464	12	US-10-161-051-119	Sequence 119, App
24	49.5	27.8	804	12	US-10-236-055A-28	Sequence 28, Appl
25	49.5	27.8	2654	15	US-10-227-610-2	Sequence 2, Appl
26	49	27.5	201	15	US-10-043-487-323	Sequence 323, App
27	49	27.5	341	12	US-09-244-805-61	Sequence 61, Appl
28	49	27.5	461	15	US-10-234-432-24	Sequence 24, Appl
29	49	27.5	524	15	US-10-253-007-56	Sequence 56, Appl
30	49	27.5	535	15	US-10-043-487-306	Sequence 306, App
31	49	27.5	893	15	US-10-234-432-86	Sequence 86, Appl
32	49	27.5	1413	12	US-09-840-743-8	Sequence 8, Appl
33	49	27.5	2000	13	US-10-010-901-29	Sequence 29, Appl
34	48.5	27.2	898	15	US-10-043-487-277	Sequence 277, App
35	48	27.0	170	9	US-09-916-790-15	Sequence 15, Appl
36	48	27.0	182	15	US-10-156-761-14821	Sequence 14821, A
37	48	27.0	485	12	US-10-287-218-20	Sequence 20, Appl
38	47.5	26.7	861	9	US-09-815-242-11081	Sequence 11081, A
39	47	26.4	529	15	US-10-156-761-10420	Sequence 10420, A
40	47	26.4	670	11	US-09-298-523B-63	Sequence 63, Appl
41	47	26.4	711	11	US-09-298-523B-3	Sequence 3, Appl
42	47	26.4	810	15	US-10-128-714-8235	Sequence 8235, Ap
43	47	26.4	811	15	US-10-128-714-3235	Sequence 3235, Ap
44	47	26.4	2462	11	US-09-819-104A-5	Sequence 5, Appl
45	46.5	26.1	100	9	US-09-867-550-2062	Sequence 2062, Ap

## ALIGNMENTS

### RESULT 1

US-10-322-579-5  
; Sequence 5, Application US/10322579  
; Publication No. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; PRIOR FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Human lgs/bcl9  
US-10-322-579-5

Query Match 100.0%; Score 178; DB 15; Length 35;  
Best Local Similarity 100.0%; Pred. No. 5e-17; Indels 0; Gaps 0;  
Matches 35; Conservative 0; Mismatches 0;  
QV 1 DGLSQEQLHRSLSQTLRDIQRLFPDEKEFTGA 35  
Db 1 DGLSQEQLHRSLSQTLRDIQRLFPDEKEFTGA 35

RESULT 2  
US-10-322-579-15

; Sequence 15, Application US/10322579  
; Publication No. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 1426  
; TYPE: PRT  
; ORGANISM: Human lgs/bcl9  
US-10-322-579-15

Query Match 100.0%; Score 178; DB 15; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 3.4e-15; Mismatches 0; Indels 0; Gaps 0;  
Matches 35; Conservative 0;

QY 1 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 35  
|||||  
DB 349 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 383

## RESULT 3

US-10-322-579-17  
; Sequence 17, Application US/10322579  
; Publication No. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 1115  
; TYPE: PRT  
; ORGANISM: Human lgs-1  
US-10-322-579-17

Query Match 60.1%; Score 107; DB 15; Length 1115;  
Best Local Similarity 87.5%; Pred. No. 8.9e-06; Mismatches 3; Indels 0; Gaps 0;  
Matches 21; Conservative 0;

QY 2 GLSQEQLHRRSLQTLRDIQRL 25  
|||||  
DB 76 GLSQEQLHRRSLQTLRDIQRL 99

## RESULT 4

US-10-322-579-4  
; Sequence 4, Application US/10322579

; Publication No. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Drosophila lgs  
US-10-322-579-4

Query Match 37.1%; Score 66; DB 15; Length 35;  
Best Local Similarity 31.4%; Pred. No. 0.054; Mismatches 10; Indels 0; Gaps 0;  
Matches 11; Conservative 10;

QY 1 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 35  
|||  
DB 1 ENLTPQQRHREQLAKIKKNQFLPENENSUGA 35

## RESULT 5

US-10-128-714-3511  
; Sequence 3511, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Broshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3511  
; LENGTH: 603  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3511

Query Match 32.3%; Score 57.5; DB 15; Length 603;  
Best Local Similarity 31.4%; Pred. No. 19; Mismatches 11; Indels 3; Gaps 1;  
Matches 11; Conservative 11;

QY 1 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 35  
|||

**Qy** 2 GLSQEQLHRRERSLQTLRD IQRM 24  
||| | | : ||| : ||| :  
**Dd** 224 GLRTEGLFRRSASVOTREIQRL 246

```

RESULT 9
US-10-097-340-21
Sequence 21, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: Manjula GANNAVAPU
APPLICANT: Sebastian HORSCH
APPLICANT: Subhangti KAWATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Michael E. MEYERS
APPLICANT: Michael MORRISEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen ZHANG
TITLE OF INVENTION: Nucleic Acid Molecules
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102

```

; SEQ ID NO 31

255 GLRTEGLFRRSASVQTVREIQRL

255 GLRTEGLFRRSASVQTVREIQRL



```

RESULT 13
US-10-097-340-16
; Sequence 16, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-16

Query Match 30.9%; Score 55; DB 15; Length 643;
Best Local Similarity 52.2%; Pred. No. 45;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSQTLRIQRM 24
DB 434 GLRTEGLFRRSASVQTVREIQL 456

RESULT 14
US-10-097-340-19
; Sequence 19, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-16

Query Match 30.9%; Score 55; DB 15; Length 643;
Best Local Similarity 52.2%; Pred. No. 45;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSQTLRIQRM 24
DB 434 GLRTEGLFRRSASVQTVREIQL 456

RESULT 15
US-10-097-340-18
; Sequence 18, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 718
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-19

Query Match 30.9%; Score 55; DB 15; Length 718;
Best Local Similarity 52.2%; Pred. No. 52;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSQTLRIQRM 24
DB 509 GLRTEGLFRRSASVQTVREIQL 531

RESULT 15
US-10-097-340-18
; Sequence 18, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 718
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-19

```

```

; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(751)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-097-340-18

```

```

Query Match      30.9%; Score 55; DB 15; Length 751;
Best Local Similarity 52.2%; Pred No. 54;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      2 GLSQQLHRSLSQTLRDIQRM 24
DB      509 GLRTEGLFRSASVQTVREIQRL 531

```

Search completed: November 13, 2003, 09:39:45  
Job time : 9.39689 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 4.76654 Seconds  
(without alignments)  
706.153 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178  
Sequence: 1 DGSQEQLEHRSLSQTLRIQRMFPDEKFTGA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76: \*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	32.0	1034	AB0551	exonuclease SbcC [imported] - Salmonella enterica subsp. enterica serovar Typhi (stra
2	56	31.5	584	S51882	topoisomerase I-re
3	56	31.5	859	T46372	hypothetical prote
4	55	30.9	287	AE2895	transcription regu
5	55	30.9	295	H97670	hex regulon repres
6	55	30.9	643	B59436	Rho GTPase activat
7	54.5	30.6	1171	E14085	phytochrome B xi
8	53.5	30.1	4957	T03455	ALR protein - huma
9	53.5	30.1	5262	T03454	hypothetical prote
10	53	29.8	227	B70438	gamma-D-glutamyl-L
11	53	29.8	376	E69957	mismatch DNA recog
12	53	29.8	818	S62790	hypothetical prote
13	52.5	29.5	242	B70356	hypothetical prote
14	52.5	29.5	705	T24343	hypothetical prote
15	52.5	29.5	1039	T14802	phytochrome B - so
16	52	29.2	332	B7017	probable transcrip
17	52	29.2	332	AD2541	transcription init
18	52	29.2	572	D92984	pyruvate dehydroge
19	52	29.2	1009	S61174	hypothetical prote
20	51	28.7	102	AH0216	conserved hypotet
21	51	28.7	237	A49940	probable alpha hel
22	51	28.7	237	E31086	probable alpha hel
23	51	28.7	237	A85901	probable alpha hel
24	51	28.7	329	D96834	hypothetical prote
25	51	28.7	477	T18801	hypothetical prote
26	51	28.7	899	GNVVM	pol polyprotein -
27	51	28.7	1047	G90684	ATP-dependent dsDN
28	51	28.7	1047	C85535	ATP-dependent dsDN
29	51	28.7	1464	S58984	development protei

## ALIGNMENTS

## RESULT 1

AB0551  
exonuclease SbcC [imported] - Salmonella enterica subsp. enterica serovar Typhi (stra  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB0551  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AB0551  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1034 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08850.1; PID:gi16501663; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY0429  
C:Superfamily: sbcC protein

Query Match 32.0%; Score S7: DB 2; Length 1034;  
Best Local Similarity 42.9%; Pred. No. 26;  
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQEQLEHRSLSQTLRIQRMFPDEK 30  
DB 213 LADEQLQLEASLQALTDEKRLADQQ 240

## RESULT 2

S51882  
topoisomerase I-related protein TRF4 - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein HRC584; protein O0716; protein YOL115W  
C:Species: Saccharomyces cerevisiae  
C:Date: 05-May-1995 #sequence\_revision 03-Aug-1995 #text\_change 21-Jul-2000  
C:Accession: S51882; S59159; S58774; S68811  
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
submitted to the EMBL Data Library, January 1995  
A:Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV includi  
and a Delta.  
A:Reference number: S51848  
A:Accession: S51882  
A:Molecule type: DNA  
A:Residues: 1-584 <VAN>  
A:Cross-references: EMBL:Z48149; NID:g663234; PID:g663237  
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
Yeast 11, 1069-1075, 1995  
A:Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including th  
a delta element.

endopeptidase Clp  
conserved hypotet  
hypothetical prote  
hypothetical prote  
CDS protein F9L11.  
hypothetical prote  
conserved hypotet  
hypothetical prote  
1-pyrroline-5-carb  
troponin T, cardia  
hypothetical prote  
hypothetical prote  
L-lactate dehydrog  
glycerol-3-phospha  
conserved hypotet  
hypothetical prote

30 50.5 28.4 835 2 AD3441  
31 50 28.1 273 2 H69337  
32 50 28.1 275 2 H69843  
33 50 28.1 319 2 S49771  
34 50 28.1 518 2 G86454  
35 50 28.1 788 2 S67595  
36 50 28.1 1162 2 D83454  
37 50 28.1 1236 2 B36329  
38 49.5 27.8 266 2 AB1124  
39 49.5 27.8 302 1 TPCHTC  
40 49.5 27.8 336 2 S72858  
41 49.5 27.8 830 2 T18860  
42 49 27.5 310 2 A84142  
43 49 27.5 338 2 E81730  
44 49 27.5 420 2 T41236  
45 49 27.5 537 2 C90796

A:Reference number: S59156; MUID:96076631; PMID:7502582  
 A:Accession: S59158  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-584 <VAV>  
 A:Cross-references: EMBL:248149; NID:9663234; PIDN:CAA88145.1; PID:9663237  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995  
 R:Sadoff, B.U.; Heath-Pagliuso, S.; Castano, I.B.; Zhu, Y.; Kieff, F.S.; Christman, M.F.  
 Genetics 141, 465-479, 1995  
 A:Title: Isolation of mutants of *Saccharomyces cerevisiae* requiring DNA topoisomerase I.  
 A:Reference number: S58774; MUID:96109595; PMID:8647385  
 A:Accession: S58774  
 A:Molecule type: DNA  
 A:Residues: 1-584 <SAD>  
 A:Cross-references: EMBL:U31355; NID:9950225; PIDN:AAC49091.1; PID:9950226  
 R:Ducand, P.; Wilger, F.; Portetelle, D.; Vandenbol, M.  
 submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S66791  
 A:Accession: S66811  
 A:Molecule type: DNA  
 A:Residues: 1-584 <DUR>  
 A:Cross-references: EMBL:Z74857; NID:91419986; PID:e251905; PID:gl1419987; MIPS:YOL115W  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD,TRP4  
 A:Cross-references: SGD:S0005475; MIPS:YOL115W  
 A:Map position: 15L  
 C:Keywords: nucleus

Query Match 31.5%; Score 56; DB 2; Length 584;  
 Best Local Similarity 36.0%; Pred. No. 19;  
 Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 4 SQQLHRSLSQTLRDQRMFLPD 28  
 DB 193 SREIEIRNQITSTIREAVKQLWPD 217

## RESULT 3

T46372  
 hypothetical protein DKFZp434p181.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46372  
 R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23031  
 A:Accession: T46372  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-859 <AAA>  
 A:Cross-references: EMBL:AL137528  
 A:Experimental source: adult testis; clone DKFZp434p1818  
 C:Genetics:  
 A:Note: DKFZp434p181.1

Query Match 31.5%; Score 56; DB 2; Length 859;  
 Best Local Similarity 33.3%; Pred. No. 29;  
 Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLEHRSLSQTLRDQRMFLPDDEKFTG 34  
 DB 317 ENQSHQELISQLQSYMKLLLPDDEKFTG 346

## RESULT 4

AE2895  
 transcription regulator, RpiR family Atu2598 [imported] - *Agrobacterium tumefaciens* (str  
 C:Species: *Agrobacterium tumefaciens*  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AE2895  
 R:Wood, D.W.; Scubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 Prage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
 ster, E.M.  
 A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AB2895  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-287 <KUR>  
 A:Cross-references: GB:AS008688; PIDN:AAL43579.1; PID:gl17741095; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu2598  
 A:Map position: circular chromosome  
 C:Superfamily: hypothetical protein ybbH

Query Match 30.9%; Score 55; DB 2; Length 287;  
 Best Local Similarity 40.0%; Pred. No. 11;  
 Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

QY 9 EHRERSLSQTLRDQRMFLP-----PDEKFTG 34  
 DB 257 QQRQSMVTLRHKQQLVEHRDPDDKQLLG 286

## RESULT 5

H97670  
 hex regulon repressor [imported] - *Agrobacterium tumefaciens* (strain C58, Cerson)  
 C:Species: *Agrobacterium tumefaciens*  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C:Accession: H97670  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldm  
 Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*  
 A:Reference number: A97359; MUID:21608551; PMID:11743194  
 A:Accession: H97670  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-295 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK88321.1; PID:gl15157797; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_4707  
 A:Map position: circular chromosome  
 C:Superfamily: hypothetical protein ybbH

Query Match 30.9%; Score 55; DB 2; Length 295;  
 Best Local Similarity 40.0%; Pred. No. 12;  
 Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

QY 9 EHRERSLSQTLRDQRMFLP-----PDEKFTG 34  
 DB 265 QQRQSMVTLRHKQQLVEHRDPDDKQLLG 294

## RESULT 6

B59436  
 Rho GTPase activating protein RhoGAP8 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 23-Sep-2002  
 C:Accession: B59436  
 R:Goward, M.E.; Huckle, E.J.  
 submitted to GenBank, April 2000  
 A:Reference number: B59436  
 A:Accession: B59436  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-643 <GOW>  
 A:Cross-references: GB:CA890248; PID:g7711011; PIDN:CA890248.1

Query Match 30.9%; Score 55; DB 2; Length 643;  
 Best Local Similarity 52.2%; Pred. No. 28;

[illegible]



Thu Nov 13 10:41:04 2003

us-09-915-543-15\_copy\_349\_383.rpr

Page 5

Search completed: November 13, 2003, 09:29:17  
Job time : 7.76654 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 3.1323 Seconds  
(without alignments)  
525.472 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178  
Sequence: 1 DGLSQQLHREHRSICLTLDIORMLPDDEKFTGA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	178	100.0	1426	1 BCL9_HUMAN	Q00512 homo sapien
2	66	37.1	1469	1 BCL9_DROME	Q961d9 drosophila
3	56	31.5	584	1 TRF4_YEAST	P53632 saccharomyc
4	55	30.9	426	1 HEMI_SYNEL	Q8d153 synchocococ
5	55	30.9	718	1 RHGB_HUMAN	Q9nsg0 homo sapien
6	54.5	30.6	174	1 ASH3_MOUSE	Q911r7 mus musculu
7	54.5	30.6	1171	1 PHYB_ORYSA	P25764 oryza sativ
8	53	29.8	375	1 YQGT_BACSU	P54497 bacillus su
9	53	29.8	817	1 MUTS_THECA	Q921x6 thermus cal
10	53	29.8	818	1 MUTS_THETH	Q56239 thermus the
11	52.5	29.5	180	1 ASH3_HUMAN	Q9nq33 homo sapien
12	52.5	29.5	190	1 RRF_FUSNN	Q8r529 fuscobacteri
13	52.5	29.5	242	1 Y755_AQUAE	O66957 aquifex aeo
14	52.5	29.5	1178	1 PHYB_SORBI	P93527 sorghum bic
15	52	29.2	332	1 RPSB_ANASP	Q03865 anabaena ep
16	52	29.2	811	1 MUTS_THEAP	Q56215 thermus aqu
17	52	29.2	1009	1 RGA2_YEAST	Q06407 saccharomyc
18	51	28.7	237	1 YFHG_ECOLI	P37328 escherichia
19	51	28.7	899	1 POL_MMTVB	P03365 mouse mamma
20	50	28.1	425	1 RHGB_MOUSE	Q9cpx4 mus musculu
21	50	28.1	2004	1 MOZ_HUMAN	Q92794 homo sapien
22	49.5	27.8	301	1 TRT2_CHICK	Q02642 gallus gall
23	49.5	27.8	830	1 DYN_CAEEL	P39055 caenorhabdi
24	49	27.5	310	1 LKH_BACHD	Q9K528 bacillus ha
25	49	27.5	328	1 PLSX_STAAM	Q8nx10 staphylococ
26	49	27.5	661	1 YDBJ_SCHPO	Q10369 schizosacch
27	49	27.5	1207	1 DML1_ARATH	Q9s1g6 arabidopsis
28	49	27.5	1234	1 PTP3_HUMAN	Q01970 homo sapien
29	48.5	27.2	184	1 YWKC_BACSU	P46870 bacillus su
30	48.5	27.2	1073	1 CARB_METWA	P58944 methanosarc
31	48.5	27.2	1132	1 PHYB_TOBAC	P29130 nicotiana t
32	48.5	27.2	1505	1 SCF2_RAT	Q70608 rattus norv
33	48.5	27.2	1872	1 T2D1_HUMAN	P21675 homo sapien

34	48	27.0	328	1 PLSX_STAAM	Q99un9 staphylococ
35	48	27.0	379	1 HS7X_PIG	P34934 sus scrofa
36	48	27.0	408	1 BTN1_YEAST	P47040 saccharomyc
37	48	27.0	650	1 CC45_YEAST	Q08032 saccharomyc
38	48	27.0	664	1 NTP1_ENTHR	P43439 enterococcu
39	48	27.0	664	1 PD13_SHEEP	Q02849 ovis aries
40	47.5	26.7	309	1 PYR8_METKA	Q8tvs2 methanopyru
41	47.5	26.7	558	1 MNT2_YEAST	P53059 saccharomyc
42	47.5	26.7	861	1 MUTS_HABIN	P44834 haemophilus
43	47	26.4	184	1 RRF_THETN	Q8ra24 thermotacari
44	47	26.4	247	1 HIS4_AGRIS	P58790 agrobacteri
45	47	26.4	296	1 HSLO_BACST	Q9f984 bacillus st

## ALIGNMENTS

RESULT 1					
BCL9_HUMAN					
ID_BCL9_HUMAN	STANDARD;	PRT;	1426 AA.		
AC	Q00512;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).				
GN	BCL9				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
GN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=98159621; PubMed=9490669;				
RA	Willis T.G., Zaiberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,				
RA	Jadav D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.N.,				
RA	Dyer M.J.S.;				
RT	"Molecular cloning of translocation t(1;14) (q21;q32) defines a novel				
RT	gene (BCL9) at chromosome 1q21.";				
RL	Blood 91:1873-1881(1998).				
RN	[2]				
RP	FUNCTION.				
RX	MEDLINE=21952490; PubMed=11955446;				
RA	Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,				
RA	Murone M., Zwelling S., Bagler K.;				
RT	"Wnt/wingless signaling requires BCL9/legless-mediated recruitment of				
RT	pypopus to the nuclear beta-catenin-TCF complex.";				
RL	Cell 109:47-60(2002).				
CC	-!- FUNCTION: Involved in signal transduction through the wnt pathway.				
CC	-!- SUBUNIT: Binds to beta-catenin (CTNBL1), PYGO1 and PYGO2.				
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	-!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,				
CC	testis, ovary and small intestine, and at lower levels in spleen,				
CC	colon and blood.				
CC	-!- DISEASE: Involved in a t(1;14) (q21;q32) chromosomal translocation				
CC	found in a patient with precursor B-cell acute lymphoblastic				
CC	leukemia (ALL). This translocation leaves the coding region				
CC	intact, but may have pathogenic effects due to alterations in the				
CC	expression level of BCL9. Several cases of translocations within				
CC	the 3' untranslated region of BCL9 have been found in B-cell				
CC	malignancies.				
CC	-!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.				
CC	-!- CAUTION: Ref.1 sequence differs from that shown due to a				
CC	frameshift in position 1391.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				



DR EMBL; Y13620; CAA73942.1; ALT\_FRAME.  
 DR Genew; HGNC:1008; BCL9.  
 DR MIM; 602597; -.  
 DR CO; GO:0007048; P:oncogenesis; TAS.  
 KW Nuclear protein; Chromosomal translocation; Proto-oncogene;  
 Wnt signaling pathway.  
 FT DOMAIN 231 1378  
 FT CTRNBL-BINDING.  
 FT DOMAIN 347 377  
 FT POLY-PRO 1.  
 FT DOMAIN 331 335  
 FT POLY-PRO 2.  
 FT DOMAIN 514 517  
 FT POLY-ALA.  
 FT DOMAIN 900 903  
 FT POLY-PRO 3.  
 FT DOMAIN 970 973  
 FT POLY-PRO 4.  
 SQ SEQUENCE 1426 AA; 149314 MW; A240A48716B7F1B CRC64;  
 Query Match 100.0%; Score 178; DB 1; Length 1426;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-15;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGLSQQLHRRSRLQTLRDIQRLFPDEKEFTGA 35  
 DB 349 DGLSQQLHRRSRLQTLRDIQRLFPDEKEFTGA 383  
 RESULT 2  
 ID BCL9\_DROME STANDARD; PRT; 1469 AA.  
 AC Q961D9; Q9V4D2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bcl-9 homolog (Legless protein).  
 GN BCL9 OR LGS OR CG2041.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Mecoptera; Pterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballwe R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen C.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Scapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wesserman D.A., Weinstein G.N., Weissbach J.,

Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of Drosophila melanogaster";  
 Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Embryo;  
 MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 "A Drosophila full-length cDNA resource";  
 Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RN [3]  
 RP SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND  
 ILE-537.  
 RC MEDLINE=21952490; PubMed=11955446;  
 RA Kramps T., Peter O., Brunner E., Neellen D., Froesch B., Chatterjee S.,  
 RA Marone M., Zuellig S., Basler K.;  
 "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
 pygopus to the nuclear beta-catenin-TCF complex";  
 Cell 109:47-60(2002).  
 RT Cell 109:47-60(2002).  
 CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.  
 CC -!- SUBUNIT: Binds to ARM and PYGO.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
 throughout development.  
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 gene model prediction.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to a license@isb-sib.ch).  
 DR EMBL; AE003844; AAF59345.1; ALT\_SEQ.  
 DR EMBL; AY051851; AAX93075.1; -.  
 DR EMBL; AF457205; AAL91368.1; -.  
 DR FlyBase; FBgn0039907; lgs.  
 DR GO; GO:0005634; C:nucleus; IEPI.  
 DR GO; GO:0005288; P:transcription regulator activity; IPI.  
 DR GO; GO:0030177; P:positive regulation of Wnt receptor signaling; IPI.  
 DR GO; GO:0007367; P:segment polarity determination; IMP.  
 KW Nuclear protein; Developmental protein; Segmentation polarity protein;  
 Wnt signaling pathway.  
 FT DOMAIN 511 555  
 FT ASB-BINDING.  
 FT DOMAIN 1134 1173  
 FT ASN-RICH.  
 FT DOMAIN 1340 1449  
 FT GLN-RICH.  
 FT DOMAIN 1162 1169  
 FT POLY-ASN.  
 FT MUTAGEN 514 514  
 FT G->E; IN ALLELE LGS-21L.  
 FT MUTAGEN 534 534  
 FT L->F; IN ALLELE LGS-17E; SEGMENT POLARITY  
 PHENOTYPE.  
 FT MUTAGEN 537 537  
 FT I->K; IN ALLELE LGS-17P.  
 SQ SEQUENCE 1469 AA; 153759 MW; 5672E01B7200ED08 CRC64;  
 Query Match 37.1%; Score 66; DB 1; Length 1469;  
 Best Local Similarity 31.4%; Pred. No. 1.2;  
 Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 DGLSQQLHRRSRLQTLRDIQRLFPDEKEFTGA 35  
 DB 520 ENITPQQRHRESEQLAKIKKQWFLFPENSVGA 554  
 RESULT 3  
 TRF4 YEAST  
 ID TRF4 YEAST PRT; 584 AA.  
 AC P53632;

01-OCT-1986 (Rel. 34, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
TOPOISOMERASE I-related protein TRF4.  
TRF4 OR YOL115W OR O0716 OR HRC584.  
Saccharomyces cerevisiae (Baker's Yeast).  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
NCBI\_TaxID=4932;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=96109595; PubMed=8647385;  
Sadoff B.U., Heath-Pagliuso S., Castano I.B., Zhu Y., Kieff P.S.,  
Christman M.F.;  
"Isolation of mutants of Saccharomyces cerevisiae requiring DNA  
topoisomerase I";  
Genetics 141:465-479(1995).  
[2]  
SEQUENCE FROM N.A.  
MEDLINE=96076631; PubMed=7502582;  
Vandenbol M., Durand P., Portetelle D., Hilger F.;  
"Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV  
including the TyI-H3 retrotransposon, the sufl(+), frameshift  
suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a  
delta element";  
Yeast 11:1069-1075(1995).  
RL FUNCTION: ESSENTIAL PROTEIN REQUIRED FOR PROPER NUCLEAR DIVISION  
CC IN MITOSIS. MAY MEDIATE MITOTIC CHROMOSOME CONDENSATION.  
CC -!- SIMILARITY: BELONGS TO THE CIDI/TRF4/TRF5 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U31355; AAC49091.1; .  
CC EMBL; Z48149; CA88145.1; .  
CC EMBL; Z74857; CA99134.1; .  
CC PIR; S51882; S51382.  
CC SGD; S0005475; TRF4.  
CC GO; GO:0005634; C:nucleus; IDA.  
CC GO; GO:0003887; F:DNA-directed DNA polymerase activity; IDA.  
CC GO; GO:0006265; P:DNA topological change; IGI.  
CC GO; GO:0007076; P:mitotic chromosome condensation; IMP.  
CC InterPro; IPR002934; NTP transf.  
CC InterPro; IPR001201; PAP\_25A core.  
CC InterPro; IPR002058; PAP\_assoc.  
CC Pfam; PF01909; NTP\_transf\_2; 1.  
CC Pfam; PF03828; PAP\_assoc; 1.  
CC Mitosis.  
KW SEQUENCE 584 AA; 66030 MW; 8AS8B29E4BFDC022 CRC64;  
SQ  
  
Query Match 31.5%; Score 56; DB 1; Length 584;  
Best Local Similarity 36.0%; Pred. No. 8.7;  
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;  
  
QY 4 SQEQLEHRSLSQTLDIQMLFPD 28  
|::||::||::||::||::||:  
DB 193 SREEIEIRNQTISTREAVKQLWPD 217  
  
RESULT 4  
HEMI SYNEL STANDARD; PRT; 426 AA.  
AC QED153.  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
GLutamyl-tRNA reductase (EC 1.2.1.-) (Glutr).  
GN HEMA OR TLL1738.  
  
OS Synechococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimpou S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
"Complete genome structure of the thermophilic cyanobacterium  
Thermosynechococcus elongatus BP-1";  
DNA Res. 9:123-130(2002).  
RL  
CC -!- CATALYTIC ACTIVITY: GlutamyI-tRNA(Glu) + NADPH = glutamate-1-  
CC semialdehyde + NADP(+); + tRNA(Glu).  
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.  
CC Involved in chlorophyll biosynthesis.  
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AP005375; BAC09290.1; .  
CC HAMAP; MF\_00087; .  
CC InterPro; IPR000343; Glutr.  
CC InterPro; IPR000594; Thlp domain.  
CC Pfam; PF00745; Glutr\_dimer; 1.  
CC Pfam; PF05201; Glutr\_M; 1.  
CC Pfam; PF05200; Glutr\_NAD\_bind; 1.  
CC TIGRfam; TIGR01035; hema; 1.  
CC PROSITE; PS00747; GLUTR; FALSE NEG.  
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;  
KW NADP; Complete proteome.  
FT ACT\_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).  
FT ACT\_SITE 99 99 BASE (BY SIMILARITY).  
SQ SEQUENCE 426 AA; 47596 MW; D84CE5A1D2AA77E CRC64;  
  
Query Match 30.9%; Score 55; DB 1; Length 426;  
Best Local Similarity 55.6%; Pred. No. 8.2;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 SQEQLEHRSLSQTLDRI 21  
|::||::||::||::||:  
DB 401 SQEDLESQRATLTQDL 418  
  
RESULT 5  
RHGS HUMAN STANDARD; PRT; 718 AA.  
ID RHGS\_HUMAN  
AC Q9NSG0; Q75983; Q95695; Q96RW1; Q96RW2; Q9HA49; Q9HC46; Q9NVX8;  
AC Q9NXL1; Q9UH20;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Rho-GTPase-activating protein 8.  
GN ARHGAP8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 4).  
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,  
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,  
RA Yu J., Han L.H.;  
"Novel human cDNA clone with function of inhibiting cancer cell  
growth".

Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

12) SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA McDermaid H.E., Hu S., Grundy P., Trichet V.;  
 RT ARHGAP8: a putative tumor-suppressor gene on chromosome 22q13.3;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 5; 6 AND 7).  
 RC TISSUE=Colon mucosa, and Mammary gland;  
 RA Isogai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hozoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT "NED0 human cDNA sequencing project";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RA Goward M.E., Huckle E.J.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Copley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Evans P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Dhani K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
 RA Hunt R.E., Hall-Tamlyn C., Heathcote R.W., Ho S., Holmes S.,  
 RA Hall S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.B., Langford C.F., Leverisha M.A., Lloyd C., Lloyd P.M.,  
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA McElay J., McLaren S., McMuray A.A., Milne S.A., Mortimore B.J.,  
 RA Odell C.N., Pavitt R., Pearson A.V., Pearson S., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lal H., Leo H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malej E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Rindes K., Kemp K., Latreille P., Layman D., Ozerisky P., Rohlfing T.,  
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,  
 RA McDermaid H.E., Johnston A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Peyvard M., Kedra D.,  
 RA Seroussi E., Fransoni I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tilahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 CC -!- FUNCTION: GTPase activator for the Rho-type GTPases by converting  
 CC them to an inactive GDP-bound state (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=8;  
 CC Comment=Additional isoforms seem to exist. Full isoforms so far

CC detected are isoforms 1 to 7. Experimental confirmation may be  
 CC lacking for some isoforms;  
 CC Name=8;  
 CC IsoId=Q9NSG0-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=Q9NSG0-2; Sequence=VSP\_001645, VSP\_001649, VSP\_001652;  
 CC Name=2;  
 CC IsoId=Q9NSG0-3; Sequence=VSP\_001645, VSP\_001649, VSP\_001653,  
 CC VSP\_001655;  
 CC Name=3;  
 CC IsoId=Q9NSG0-4; Sequence=VSP\_001651;  
 CC Name=4;  
 CC IsoId=Q9NSG0-5; Sequence=VSP\_001650, VSP\_001654;  
 CC Name=5;  
 CC IsoId=Q9NSG0-6; Sequence=VSP\_001647;  
 CC Name=6;  
 CC IsoId=Q9NSG0-7; Sequence=VSP\_001647, VSP\_001652, VSP\_001656,  
 CC VSP\_001657;  
 CC Name=7;  
 CC IsoId=Q9NSG0-8; Sequence=VSP\_001646, VSP\_001648, VSP\_001650,  
 CC VSP\_001654;  
 CC -!- SIMILARITY: Contains 1 CRAL-TRIO domain.  
 CC -!- SIMILARITY: Contains 1 Rho-GAP domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF177331; AAG17975.1; -;  
 CC EMBL; AF195968; AAK58136.1; -;  
 CC EMBL; AF195965; AAK58137.1; -;  
 CC EMBL; AK000192; BAA90999.1; -;  
 CC EMBL; AK001306; BAA91614.1; -;  
 CC EMBL; AK022305; BAB14008.1; -;  
 CC EMBL; AL355192; CAB90248.1; -;  
 CC EMBL; 298743; CAB1416.1; ALT\_INIT.  
 CC EMBL; 293244; CAB07531.1; ALT\_SEQ.  
 CC EMBL; 283838; CAB62993.1; ALT\_INIT.  
 CC FIR; B59436; B59436.  
 CC HSSP; Q07960; IRGP.  
 CC Genew; HGNC:677; ARHGAP8.  
 CC InterPro; IPR001251; CRAL TRIO.  
 CC InterPro; IPR00198; RhoGAP.  
 CC Pfam; PF00620; RhoGAP; 1.  
 CC SMART; SM00124; RhoGAP; 1.  
 CC PROSITE; PS50191; CRAL TRIO; 1.  
 CC PROSITE; PS50238; RhoGAP; 1.  
 CC GTPase activation; Alternative splicing.  
 CC DOMAIN 267 453  
 CC DOMAIN 490 666  
 CC VARSPLIC 1 95  
 CC Missing (in isoform 1 and isoform 2).  
 CC VARSPLIC 1 101  
 CC /FtId=VSP\_001645.  
 CC Missing (in isoform 7).  
 CC /FtId=VSP\_001646.  
 CC VARSPLIC 1 254  
 CC Missing (in isoform 5 and isoform 6).  
 CC /FtId=VSP\_001647.  
 CC KIRPVE -> MAPWPT (in isoform 7).  
 CC /FtId=VSP\_001648.  
 CC Missing (in isoform 1 and isoform 2).  
 CC /FtId=VSP\_001649.  
 CC LQDRAAAAVLGAARRKRRPSVPMAGQDPLSTSHPYDVA  
 CC RHGLQVAGDRFRVVTSCCRPPSHLDHQRLLLEYLK  
 CC YTLQYVENDYITFYHYGLNSRNKPSLGLQSAKFEFRK  
 CC DGLDTPMRLVNSKLRSSHLSPKYWDYRYKK -> KRL  
 CC LRRSRGDVLAKNPVRSKSYNTPLNPVQHEAEGAAAGG  
 CC TSIRHVSVEVSTCEPQGFSDPPCGPTGTGTSSPAPHS  
 CC PCSELVPTTQPPQGLDTPRSSLPSPSPENLDQLLESVD  
 CC SDSEIFIDFGRGSGMSDLGSGGQSVV (in  
 CC isoform 4 and isoform 7).

```

FT FT VARSPLIC 311 385 /FTid=VSP_001650.
FT FT Missing (in isoform 3).
FT FT /FTid=VSP_001651.
FT FT VARSPLIC 355 385 Missing (in isoform 1 and isoform 6).
FT FT /FTid=VSP_001652.
FT FT VARSPLIC 386 451 YKGLKALIVVFTSIKIVLRLYDEK -> QEPQCQANTLVL
FT FT LSELHLHLYDQVIVPEVLRVDEK -> QEPQCQANTLVL
FT FT KGPDSQHFSFAGLLYCNNSAGLCSKTLWTLKCEPHVIFT
FT FT CHEIFFCFPFST (in isoform 2).
FT FT /FTid=VSP_001653.
FT FT VARSPLIC 389 718 Missing (in isoform 4 and isoform 7).
FT FT /FTid=VSP_001654.
FT FT VARSPLIC 452 718 Missing (in isoform 2).
FT FT /FTid=VSP_001655.
FT FT VARSPLIC 579 579 V -> A (in isoform 6).
FT FT /FTid=VSP_001656.
FT FT VARSPLIC 580 718 Missing (in isoform 6).
FT FT /FTid=VSP_001657.
FT FT /FTid=VSP_001658.
SQ SEQUENCE 718 AA; 81953 MW; 7181835625487465 CRC64;

Query Match 30.9%; Score 55; DB 1; Length 718;
Best Local Similarity 52.2%; Pred. No. 15;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLEHRSRSLQTLRIQRM 24
DB 509 GLRTEGLFRSASVQTVREIQL 531

RESULT 6
ASH3_MOUSE
ID ASH3_MOUSE STANDARD; PRT; 174 AA.
AC Q9JUR7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Achaete-scute homolog 3 (bHLH transcriptional regulator Sgn-1) (Mash-3).
DE ASCL3 OR MASH3 OR SGNI.
GN Mus musculus (Mouse).
OS Mus musculus; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA Kemp P.R., Cooper W.N., Metcalfe J.C.;
RP "MASH3 a novel basic helix-loop-helix protein that inhibits myogenesis in C2C12 cells."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA STRAIN=C57BL/6;
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S., Ohno K., Takakura A., Takebayashi H., Okada T., Abe K., Nabeshima Y.;
RT "Sgn1, a basic helix-loop-helix transcription factor delineates the salivary gland duct cell lineage in mice."
RL Dev. Biol. 240:517-530(2001).
RN [3]
SEQUENCE FROM N.A.
RA MEDLINE=21418998; PubMed=11528127;
RX Amid C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A., Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar architecture of a conserved syntenic region on human chromosome 1p15.3 (including gene STS) and mouse chromosome 7."
RL Cytogenet. Cell Genet. 93:284-290(2001).
CC -!- FUNCTION: Transcriptional repressor. Inhibits myogenesis.
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another bHLH protein.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Specifically expressed in the salivary duct cells.

```

```

CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ277605; CAC37689.1; -
CC EMBL; AB046448; BAB83911.1; -
CC EMBL; AJ400878; CAB92296.1; -
CC MGD; MGI:1928820; Ascl3.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0005667; C:transcription factor complex; IPI.
CC GO; GO:0003677; F:DNA binding activity; IDA.
CC GO; GO:0005515; F:protein binding activity; IPI.
CC GO; GO:0030528; F:transcription regulator activity; IDA.
CC GO; GO:0006357; P:regulation of transcription from Pol II promoter; IDA.
CC InterPro: IPR001092; HLH_basic.
CC Pfam; PF00010; HLH; 1.
CC SMART; SM00353; HLH; 1.
CC PROSITE; PS00038; HLH 1; 1.
CC PROSITE; PS00888; HLH 2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA_BIND 95 105 BASIC DOMAIN
FT DOMAIN 106 145 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ SEQUENCE 174 AA; 20245 MW; D89E56C8A9D3440B CRC64;

Query Match 30.6%; Score 54.5; DB 1; Length 174;
Best Local Similarity 36.1%; Pred. No. 3.5;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;

QY 3 LSQQLEHRSRSLQTLRD-----IQRLPDEKE 31
DB 119 LPEDYLEKRLSKVETLRAAIKYISYQLSLYPDESE 154

RESULT 7
PHYB_ORYZA
ID PHYB_ORYZA STANDARD; PRT; 1171 AA.
AC P25764;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytochrome B.
GN PHYB OR PHYB1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=cv. Indica-IR36; TISSUE=Seedling shoot;
RX MEDLINE=91172131; PubMed=2005872;
RA Debesh K., Tepperman J., Christensen A.H., Quail P.H.;
RT "phyB is evolutionarily conserved and constitutively expressed in rice seedling shoots."
RL Mol. Gen. Genet. 225:305-313(1991).
CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN, PROTOCHLOROPHYLLIDE REDUCTASE, RRNA, ETC. IT ALSO CONTROLS THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.

```

CC -!- SUBUNIT: Homodimer.  
 CC -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) dimerization domains.  
 CC -!- SIMILARITY: CONTAINS 1 histidine kinase domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X57563; CAA0795.2; --  
 CC FIR: S14065; S14065.  
 CC GRAMME: P25764; --  
 CC InterPro: IPR003594; ATPbind\_ATPase.  
 CC InterPro: IPR003018; GAF.  
 CC InterPro: IPR003661; His\_kinA.  
 CC InterPro: IPR005467; His\_kinase.  
 CC InterPro: IPR000014; PAS\_domain.  
 CC InterPro: IPR001294; Phytochrome.  
 CC Pfam: PF01590; GAF; 1.  
 CC Pfam: PF02518; HATPase\_c; 1.  
 CC Pfam: PF00512; HSKA; 1.  
 CC Pfam: PF00989; PAS; 2.  
 CC Pfam: PF00360; phytochrome; 1.  
 CC PRINTS: PR01033; PHYTOCHROME.  
 CC SMART: SM00065; GAF; 1.  
 CC SMART: SM00388; HSKA; 1.  
 CC SMART: SM00091; PAS; 2.  
 CC TIGR: TIGR00229; sensory\_box; 2.  
 CC PROSITE: PS50109; HIS\_KIN; 1.  
 CC PROSITE: PS50112; PAS; 2.  
 CC PROSITE: PS00245; PHYTOCHROME\_1; 1.  
 CC PROSITE: PS50046; PHYTOCHROME\_2; 1.  
 CC Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 KW Repeat; Multigene family.  
 FT DOMAIN 661 732 PAS 1.  
 FT DOMAIN 795 866 PAS 2.  
 FT DOMAIN 943 1161 HISTIDINE KINASE.  
 FT DOMAIN 39 51 POLY-GLY.  
 FT BINDING 364 364 CHROMOPHORE (BY SIMILARITY).  
 FT SEQUENCE 1171 AA; 128384 MW; E8292588B769BF16 CRC64;  
 Query Match 30.6%; Score 54.5; DB 1; Length 1171;  
 Best Local Similarity 48.1%; Pred. No. 30;  
 Matches 13; Conservative 4; Mismatches 5; Indels 5; Gaps 1;  
 QY 3 LSQEQLEHRSRSLQTLRDIQRMFLPDE 29  
 DB 1019 VSQVMQLRERDLQLRDI-----PDE 1040  
 RESULT 8  
 YQGT BACSU  
 ID YQGT BACSU STANDARD; PRT; 376 AA.  
 AC P54437;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yqgt.  
 GN YQGT.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,

RA Kobayashi Y.;  
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
 RT the Bacillus subtilis genome containing the skin element and many  
 RT sporulation genes.";   
 RL Microbiology 142:3103-3111 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunat F., Ogaawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Brouillet S., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Broillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Rapoport G., Prescott A.M.,  
 RA Presecan E., Fujic P., Furnelle B., Fournelle S., Rose M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis".  
 RL Nature 390:249-256 (1997).  
 CC -!- SIMILARITY: STRONG, TO B.SPHERICUS GAMMA-D-GLUTAMYL-L-DIAMINO  
 CC ACID ENDOPEPTIDASE I (EC 3.4.99.11).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D84432; BAA12523.1; --  
 CC FIR: E69957; E69957.  
 CC MEROPS: M14.008; --  
 CC Subtilist; BG11687; yqgt.  
 CC InterPro: IPR000834; Zn\_carbOpept.  
 CC Pfam: PF00246; Zn\_carbOpept; 1.  
 CC PRINTS: PR00765; CRBOXYPTASEA.  
 CC SMART: SM00631; Zn\_pept; 1.  
 CC Hypothetical protein; Hydrolase; Metalloprotease; Zinc;  
 KW Complete proteome.  
 SQ SEQUENCE 376 AA; 43439 MW; 1075CDC124823BB4 CRC64;  
 Query Match 29.8%; Score 53; DB 1; Length 376;  
 Best Local Similarity 52.6%; Pred. No. 13;  
 Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 9 EHRERSLQTLRDIQRMFLP 27  
 DB 50 EHSKELTIQDIKRFLOP 68

## RESULT 9

ID MUTS\_THETH STANDARD; PRT; 817 AA.  
 AC Q92IX6.  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA mismatch repair protein muts.  
 GN MUTS.  
 OS Thermus caldophilus.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 OX NCBI\_TaxID=272;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GX24;  
 RA Nishiru O., Park B.C., Ko J.H., Kim J.S., Koh S.K., Lee H.C.,  
 RA Kim C.H., Lee S.K., Lee D.-S.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC FUNCTION: This protein is involved in the repair of mismatches in  
 CC DNA. It is possible that it carries out the mismatch recognition  
 CC step. This protein has a weak ATPase activity (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).

EMBL; AF007553; AAD01407.1; -;  
 HAMAP; MF 00096; -; 1.  
 InterPro; IPR005748; Muts1.  
 InterPro; IPR000432; Muts\_C.  
 InterPro; IPR002863; Muts\_N.  
 Pfam; PF01624; Muts\_I; 1.  
 Pfam; PF05188; Muts\_II; 1.  
 Pfam; PF05192; Muts\_III; 1.  
 Pfam; PF05190; Muts\_IV; 1.  
 Pfam; PF00488; Muts\_V; 1.  
 ProDom; PD001263; Muts\_C; 1.  
 SMART; SM00534; Mutsac; 1.  
 SMART; SM00533; Mutsd; 1.  
 TIGRFAMs; TIGR01070; muts1; 1.  
 PROSITE; PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
 KW DNA repair; ATP-binding; DNA-binding.  
 FT NP\_BIND 589 596 ATP (POTENTIAL).  
 SQ SEQUENCE 817 AA; 91179 MW; 61EA066FB84BA761 CRC64;

Query Match 29.8%; Score 53; DB 1; Length 817;  
 Best Local Similarity 42.3%; Pred. No. 31;  
 Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLEHRSLSQTLRDIOQLMFPDEKE 31  
 Db 464 EKVPQYRPVQTLKDRQRYTLPEMKE 489

## RESULT 10

ID MUTS\_THETH STANDARD; PRT; 818 AA.  
 AC Q56239;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA mismatch repair protein muts.  
 GN MUTS.  
 OS Thermus thermophilus.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 OX NCBI\_TaxID=274;

RP SEQUENCE FROM N.A., SEQUENCE OF 1-20, AND CHARACTERIZATION.  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=96177563; PubMed=8604304;  
 RA Takamatsu S., Kato R., Kuramitsu S.;  
 RL "Mismatch DNA recognition protein from an extremely thermophilic  
 RL bacterium, Thermus thermophilus HB8.";  
 RL Nucleic Acids Res. 24:640-647(1996).  
 [2]  
 RP DOMAIN STRUCTURE.  
 RC STRAIN=HB8 / ATCC 27634;  
 RX MEDLINE=98391738; PubMed=9722634;  
 RA Tachiki H., Kato R., Masui R., Hasegawa K., Itakura H., Fukuyama K.,  
 RA Kuramitsu S.;  
 RL "Domain organization and functional analysis of Thermus thermophilus  
 RL MutS protein.";  
 RL Nucleic Acids Res. 26:4153-4159(1998).  
 CC FUNCTION: This protein is involved in the repair of mismatches in  
 CC DNA. It is possible that it carries out the mismatch recognition  
 CC step. This protein has a maximal ATPase activity at 80 degrees  
 CC Celsius. Binds double-stranded DNA.  
 CC -!- SUBUNIT: Homotetramer.  
 CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).

EMBL; D63810; BAA09880.1; -;  
 HAMAP; MF 00096; -; 1.  
 InterPro; IPR005748; Muts1.  
 InterPro; IPR000432; Muts\_C.  
 InterPro; IPR002863; Muts\_N.  
 Pfam; PF01624; Muts\_I; 1.  
 Pfam; PF05188; Muts\_II; 1.  
 Pfam; PF05192; Muts\_III; 1.  
 Pfam; PF05190; Muts\_IV; 1.  
 Pfam; PF00488; Muts\_V; 1.  
 ProDom; PD001263; Muts\_C; 1.  
 SMART; SM00534; Mutsac; 1.  
 SMART; SM00533; Mutsd; 1.  
 TIGRFAMs; TIGR01070; muts1; 1.  
 PROSITE; PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
 KW DNA repair; ATP-binding; DNA-binding.  
 FT INIT\_MET 0 0  
 FT DOMAIN 1 129 A1.  
 FT DOMAIN 130 273 A2.  
 FT DOMAIN 274 569 B; DNA-BINDING.  
 FT DOMAIN 570 818 C.  
 FT NP\_BIND 590 597 ATP (POTENTIAL).  
 SQ SEQUENCE 818 AA; 91249 MW; 657FD676BF033506 CRC64;

Query Match 29.8%; Score 53; DB 1; Length 818;  
 Best Local Similarity 42.3%; Pred. No. 31;  
 Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLEHRSLSQTLRDIOQLMFPDEKE 31  
 Db 465 EKVPQYRPVQTLKDRQRYTLPEMKE 490

## RESULT 11

ID ASH3\_HUMAN STANDARD; PRT; 180 AA.  
 AC Q9NQ33; Q8WYQ6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Achaete-scute homolog 3 (bHLH transcriptional regulator Sgn-1).

```
GN ASCL3 OR SGNI1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S., Ohno K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.;
RT "Sgnt, a basic helix-loop-helix transcription factor delineates the
RT salivary gland duct cell lineage in mice.";
RL Dev. Biol. 240:517-530(2001).
RN SEQUENCE FROM N.A.
RX MEDLINE=21418998; PubMed=11528127;
RA Amid C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
RT architecture of a conserved syntenic region on human chromosome
RT 1p15.3 (including gene STS) and mouse chromosome 7.";
RL Cytogenet. Cell Genet. 93:284-290(2001).
CC -!- FUNCTION: Transcriptional repressor. Inhibits myogenesis (By
CC similarity).
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC bHLH protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB046450; BAB83913.1; -
CC EMBL; AJ400877; CAB92288.1; ALT_INIT.
CC Genbank; HGNC:740; ASCL3.
CC
CC InterPro: IPR001092; HLH_basic.
CC Pfam; PF00010; HLH; 1.
CC SMART; SM00353; HLH; 1.
CC PROSITE; PS00038; HLH 1; 1.
CC PROSITE; PS00888; HLH 2; 1.
CC
CC Transcription regulation; Repressor; DNA-binding; Nuclear protein.
KW DOMAIN
FT DOMAIN 95 105 BASIC DOMAIN
FT CONFLICT 106 145 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT CONFLICT 54 54 R -> L (IN REF. 1).
SQ SEQUENCE 180 AA; 20796 MW; C9B90C885DE94F81 CRC64;

Query Match 29.5%; Score 52.5; DB 1; Length 180;
Best Local Similarity 36.1%; Pred. No. 6.6;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;

QY 3 LSQQLHRRSLQTLRD-----IQRLPFDEKE 31
DB 119 LPPEYLERSLKSVETRAAIKYINLYQSLYPDKAE 154

RESULT 12
RFP_FUSNN
ID RFP_FUSNN STANDARD; PRT; 190 AA.
AC Q8R529;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RFP).
GN FRR OR FN1623.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
```

```
OX NCBI_TaxID=76856;
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Lasaine M., Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: Responsible for the release of ribosomes from messenger
CC RNA at the termination of protein biosynthesis. May increase the
CC efficiency of translation by recycling ribosomes from one round of
CC translocation to another (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RRF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB010470; AAL93738.1; -
CC HAMAP; MF 00040; -; 1.
CC InterPro: IPR002661; RRF.
CC Pfam; PF01765; RRF; 1.
CC ProDom; PD004103; RRF; 1.
CC TIGRPFAMs; TIGR00496; rrf; 1.
CC
CC Protein biosynthesis; Complete proteome.
SQ SEQUENCE 190 AA; 21438 MW; 81BD4A7A293B2814 CRC64;

Query Match 29.5%; Score 52.5; DB 1; Length 190;
Best Local Similarity 36.1%; Pred. No. 7;
Matches 13; Conservative 8; Mismatches 10; Indels 5; Gaps 1;

QY 3 LSQQLHRRSLQTLRD-----IQRLPFDEKEFT 33
DB 153 ISEDLKKEETVQTLTDKYVKEIDDLAKKEKIT 188

RESULT 13
Y755_AQUAE
ID Y755_AQUAE STANDARD; PRT; 242 AA.
AC O66957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_755.
GN AQ_755.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN SEQUENCE FROM N.A.
RX STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```





```
RT Cyano bacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001)
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED.
CC -!- SIMILARITY: Belongs to the sigma-70 factor family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95760; AAA22046.1; -.
DR EMBL; AP003602; BAB77258.1; -.
DR PIR; AD2541; AD2541.
DR HSP; P00579; ISIG.
DR InterPro; IPR000943; Sigma 70.
DR Pfam; PF00140; sigma70_x1_2; 1.
DR Pfam; PF04542; sigma70_x2; 1.
DR Pfam; PF04539; sigma70_x3; 1.
DR Pfam; PF04545; sigma70_x4; 1.
DR PRINTS; PR00046; SIGMA70FCT.
DR PROSITE; PS00715; SIGMA70_1; 1.
DR PROSITE; PS00716; SIGMA70_2; 1.
KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding; Plasmid; Complete proteome.
FT DOMAIN 125 138 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 294 313 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 332 AA; 38431 MW; 81323E670CCDF46A CRC64;

Query Match 29.2%; Score 52; DB 1; Length 332;
Best Local Similarity 35.5%; Pred. No. 15;
Matches 11; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

Qy 1 DGLSQQLHRRSLQTLRDIQRLFPDEKE 31
|||: |||: |||: |||: |||:
Db 248 DGMSPERYAERELLYQDIHNLAKLTPOQKE 278
```

Search completed: November 13, 2003, 09:25:59  
Job time : 5.1323 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 11.4397 Seconds  
(without alignments)  
789.518 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQQLERHRSRLQLRLDIQRLFPDEKEFTGA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mmc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	32.0	1034	16 Q828Y6	Q828Y6 salmonella
2	57	32.0	1046	16 Q82RE3	Q82RE3 salmonella
3	56	31.5	859	4 Q9NT51	Q9NT51 homo sapien
4	56	31.5	1132	4 Q9Y2H2	Q9Y2H2 homo sapien
5	56	31.5	1208	5 Q9VXU1	Q9VXU1 drosophila
6	56	31.5	1398	5 Q9VXU2	Q9VXU2 drosophila
7	55	30.9	295	16 Q8UCA0	Q8UCA0 agrobacteri
8	55	30.9	426	16 Q8DI53	Q8DI53 synechococc
9	55	30.9	433	4 Q8I2M6	Q8I2M6 homo sapien
10	55	30.9	818	17 Q8I2GZ1	Q8I2GZ1 methanopyru
11	53.5	30.1	155	11 Q8VD56	Q8VD56 rattus norv
12	53.5	30.1	4957	4 O14687	O14687 homo sapien
13	53.5	30.1	5262	4 O14686	O14686 homo sapien
14	53	29.8	227	16 Q67532	Q67532 aquifex seo
15	53	29.8	411	16 Q8PRD9	Q8PRD9 xanthomonas
16	52.5	29.5	673	5 O01302	O01302 caenorhabdi

17	52	29.2	399	10 Q8W0M3	Q8W0M3 oryza sativ
18	52	29.2	572	16 Q8HTQ7	Q8HTQ7 pseudomonas
19	52	29.2	616	12 P87708	P87708 fowlpox vir
20	52	29.2	946	4 Q96QE4	Q96QE4 homo sapien
21	52	29.2	1009	3 Q06407	Q06407 saccharomyc
22	52	29.2	1766	12 Q9J599	Q9J599 fowlpox vir
23	52	29.2	2607	11 Q8BTI8	Q8BTI8 mus musculu
24	51.5	28.9	1146	10 Q8FPQ2	Q8FPQ2 populus tri
25	51.5	28.9	1151	10 Q8FPQ3	Q8FPQ3 populus tri
26	51	28.7	102	16 Q8FD96	Q8FD96 yersinia pe
27	51	28.7	108	16 Q8D096	Q8D096 yersinia pe
28	51	28.7	239	16 Q8FF28	Q8FF28 escherichia
29	51	28.7	258	11 Q8BVP0	Q8BVP0 mus musculu
30	51	28.7	329	10 Q8LG66	Q8LG66 arabidopsis
31	51	28.7	329	10 Q8C975	Q8C975 arabidopsis
32	51	28.7	337	5 Q62022	Q62022 caenorhabdi
33	51	28.7	341	4 Q8I1Y3	Q8I1Y3 homo sapien
34	51	28.7	449	5 Q8I0M4	Q8I0M4 caenorhabdi
35	51	28.7	848	16 Q8FMB8	Q8FMB8 corynebacte
36	51	28.7	1047	16 Q8XEJ6	Q8XEJ6 escherichia
37	51	28.7	1464	5 Q23995	Q23995 drosophila
38	51	28.7	1464	5 Q9VC47	Q9VC47 drosophila
39	51	28.7	1484	5 Q24132	Q24132 drosophila
40	51	28.7	1755	15 Q56220	Q56220 mouse mamma
41	51	28.7	1755	15 Q9IZT3	Q9IZT3 exogenous m
42	50.5	28.4	96	5 Q9VN78	Q9VN78 drosophila
43	50.5	28.4	288	13 Q8UVZ1	Q8UVZ1 brachydanio
44	50.5	28.4	657	10 Q9FNQ6	Q9FNQ6 arabidopsis
45	50.5	28.4	675	4 Q9HD27	Q9HD27 homo sapien

# ALIGNMENTS

RESULT 1

Q828Y6	PRELIMINARY;	PRT; 1034 AA.
ID Q828Y6		
AC Q828Y6;		
DT 01-MAR-2002 (TReMBLrel. 20, Created)		
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)		
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)		
DE Exonuclease SbCC.		
GN STY0429.		
OS Salmonella typhi.		
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC Enterobacteriaceae; Salmonella.		
OX NCBI_TaxID=601;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=CT18;		
RX MEDLINE=21534947; PubMed=11677608;		
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,		
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,		
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,		
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,		
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,		
RA Kragh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,		
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,		
RA Whitehead S., Barrell B.G.;		
RT "Complete genome sequence of a multiple drug resistant Salmonella		
RL Enterica serovar Typhi CT18.";		
RL Nature 413:848-852(2001).		
DR EMBL: AL627266; CAD08850.1; "		
DR InterPro; IPR003439; ABC transporter.		
DR InterPro; IPR004592; SDC.		
DR InterPro; IPR003395; SMC.N.		
DR Pfam; PF02463; SMC.N; 1.		
DR TIGRFAMs; TIGR00618; sbcc; 1.		
KW Complete proteome.		
SQ SEQUENCE 1034 AA; 116759 MW; 2513B7573626960A CRC64;		

Query Match 32.0%; Score 57; DB 16; Length 1034;  
Best Local Similarity 42.9%; Pred. No. 68;

Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSOEQLHERSLQTLRDIQRMFLPDEK 30  
 Db 213 LADEQLQLEASLQALTDDEKRLADQ 240

RESULT 2

Q82R3 PRELIMINARY; PRT; 1046 AA.

ID Q82R3  
 AC Q82R3  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE ATP-dependent deDNA exonuclease.  
 GN SBCC OR STM0395  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 DR EMBL; AB008713; AA19349.1; --  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR004592; SbcC.  
 DR InterPro; IPR003395; SMC\_N.  
 DR Pfam; PF02463; SMC\_N; 1.  
 DR TIGRFAMs; TIGR00618; sbcc; 1.  
 DR Exonuclease; Complete proteome.  
 KW Exonuclease; Complete proteome.  
 SQ SEQUENCE 1046 AA; 117824 MW; BA565CA3BDAD0C82 CRC64;

Query Match 32.0%; Score 57; DB 16; Length 1046;  
 Best Local Similarity 42.9%; Pred. No. 69;  
 Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSOEQLHERSLQTLRDIQRMFLPDEK 30  
 Db 213 LADEQLQLEASLQALTDDEKRLADQ 240

RESULT 3

Q9NT51 PRELIMINARY; PRT; 859 AA.

ID Q9NT51  
 AC Q9NT51  
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE Hypothetical protein (fragment).  
 GN DKFZP434P1818.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL137528; CAB70792.1; --  
 DR InterPro; IPR002013; Syja\_N.  
 DR Pfam; PF02383; Syja\_N; 1.  
 DR PROSITE; PS50275; SAC; 1.  
 KW Hypothetical protein.

FT NON TER 1 1  
 SQ SEQUENCE 859 AA; 96781 MW; BCBC47C8B726D76 CRC64;

Query Match 31.5%; Score 56; DB 4; Length 859;  
 Best Local Similarity 33.3%; Pred. No. 76;  
 Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLHERSLQTLRDIQRMFLPDEKFTG 34  
 Db 317 ENQSHQELISQLQSYMKLLLPDDEKPHG 346

RESULT 4

Q9Y2H2 PRELIMINARY; PRT; 1132 AA.

ID Q9Y2H2  
 AC Q9Y2H2  
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE Hypothetical protein KIAA0966.  
 GN KIAA0966.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99246063; PubMed=10231032;  
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. XIII.  
 The complete sequences of 100 new cDNA clones from brain which code  
 for large proteins in vitro.";  
 RL DNA Res. 6:63-70(1999).  
 DR EMBL; AB023183; BAA76810.1; --  
 DR InterPro; IPR002013; Syja\_N.  
 DR Pfam; PF02383; Syja\_N; 1.  
 DR PROSITE; PS50275; SAC; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1132 AA; 128407 MW; 853719FC0AD455CD CRC64;

Query Match 31.5%; Score 56; DB 4; Length 1132;  
 Best Local Similarity 33.3%; Pred. No. 1e-02;  
 Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLHERSLQTLRDIQRMFLPDEKFTG 34  
 Db 590 ENQSHQELISQLQSYMKLLLPDDEKPHG 619

RESULT 5

Q9VXU1 PRELIMINARY; PRT; 1208 AA.

ID Q9VXU1  
 AC Q9VXU1  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE CG32587 protein.  
 GN CG32587 OR CG6350 OR CG7821.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Burtis K.C., Busam D.A., Cawley S., Dahke C., Davenport L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Floslek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Glodok A., Grogan E., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RT Science 287:2185-2195(2000).  
 RL [2]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howard T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RA "Sequencing of *Drosophila melanogaster* genome.";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochink S.E., Smith C.D.,  
 RA Tupy J.L., Berman C., Berman B., Carlson J.W., Ceiniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [5]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003499; AEF48467.2; -  
 DR FlyBase; Fg00052587; CG32587.  
 SQ SEQUENCE 1208 AA; 137774 MW; 131F2EB663826A92 CRC64;

Query Match 31.5%; Score 56; DB 5; Length 1208;  
 Best Local Similarity 45.5%; Pred. No. 1.1e+02;  
 Matches 10; Conservative 6; Indels 0; Gaps 0;  
 QY 3 LSQEQLEHRSLSQTLRDIORM 24  
 DB 811 LQQQAESEQQASTLRDLRL 832  
 ID QVXU2 PRELIMINARY; PRT; 1398 AA.  
 AC QVXU2; Q960D0.  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE CG32587 protein (SP07366P).  
 GN CG32587 OR CG6350 OR CG7821.  
 OS *Drosophila melanogaster* (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Ananides P.G., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Adams M.D., Ceiniker S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Burtis K.C., Busam D.A., Cawley S., Dahke C., Davenport L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Floslek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Glodok A., Grogan E., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RT Science 287:2185-2195(2000).  
 RL [2]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howard T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RA "Sequencing of *Drosophila melanogaster* genome.";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochink S.E., Smith C.D.,  
 RA Tupy J.L., Berman C., Berman B., Carlson J.W., Ceiniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [5]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003499; AEF48467.2; -  
 DR FlyBase; Fg00052587; CG32587.  
 SQ SEQUENCE 1208 AA; 137774 MW; 131F2EB663826A92 CRC64;

RA Perrier S., Frise E., Galle R.P., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hoslin D., Howland T.J.,  
RA Dhegwan C., Jalali M., Kuse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,  
RA Pacleb J., Paragas V., Park S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RA Masra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bergman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RN SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RN SEQUENCE FROM N.A.  
RA FlyBase;  
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RN SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,  
RA Champs M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda C.J.,  
RA Nunco J., Pacieb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RA EMBL; AE003499; AAF48466.2; -  
DR EMBL; AY052121; AAK93545.1; -  
DR FlyBase; FBGN0052587; CG32587.  
SQ SEQUENCE 1398 AA; 158484 MW; 8BA1C2FE3E9F555D CRC64;  
Query Match 31.5%; Score 56; DB 5; Length 1398;  
Best Local Similarity 45.5%; Pred. No. 1.2e+02;  
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;  
QY 3 ISQQLHRSLSLTLDIORM 24  
DB 1001 LQQQAESQQAASLTLDI 1022  
RESULT 7  
ID Q8UCAO PRELIMINARY; PRT; 295 AA.  
AC Q8UCAO;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DE Transcriptional regulator, RpiR family.  
GN ATU2598 OR AGR C.4707.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavin I., Levy R., Li M.-J., McClelland E., Palmieri A.,  
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,

RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58";  
RL Science 294:2317-2323 (2001).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mallin L.,  
RA Wolmuel K., Gordon J., Vaudin M., Lartchouk O., Epp A., Liu F.,  
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58";  
RL Science 294:2323-2328 (2001).  
DR EMBL; AE009206; AAL43579.1; ALT\_INIT.  
DR EMBL; AE008173; AAK89321.1; -  
DR InterPro; IPR000843; HTH\_Laci.  
DR InterPro; IPR000281; HTH\_RpiR.  
DR InterPro; IPR001347; SIS.  
DR Pfam; PF01418; HTH\_6; 1.  
DR Pfam; PF01380; SIS\_1.  
DR PROSITE; PS00356; HTH\_LACI\_FAMILY; 1.  
KW Complete proteome.  
SQ SEQUENCE 295 AA; 32259 MW; 3B50926B3CB72456 CRC64;  
Query Match 30.9%; Score 55; DB 16; Length 295;  
Best Local Similarity 40.0%; Pred. No. 35;  
Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 1;  
QY 9 EHRERSLQTLRDIQRMFLF----PDEKEFTG 34  
DB 265 QQRQSMVTLRHKKQLVHRDPDDKQLLG 294  
RESULT 8  
ID Q8DI53 PRELIMINARY; PRT; 426 AA.  
AC Q8DI53;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE Transfer RNA-Gln reductase.  
GN HEMA OR TIL1738.  
OS Synechococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=BP-1;  
RX MEDLINE=2225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1";  
RL DNA Res. 9:123-130(2002)  
DR EMBL; AP005375; BAC09290.1; -  
KW Complete proteome.  
SQ SEQUENCE 426 AA; 47596 MW; D84CESA1D2AA777E CRC64;  
Query Match 30.9%; Score 55; DB 16; Length 426;  
Best Local Similarity 55.6%; Pred. No. 50;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
QY 4 SQEQLEHRSLSLTLDI 21  
DB 401 SQDLESQRAMOTLQDL 418

```

RESULT 9
Q81ZM6
ID Q81ZM6 PRELIMINARY; PRT; 433 AA.
AC Q81ZM6
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE BCH domain-containing Cdc42GAP-like protein.
GN BPGAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shang X., Zhou Y.T., Low B.C.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF54240; AAN40769.1; -.
SQ SEQUENCE 433 AA; 49691 MW; 0AE4B42A404AE1D3 CRC64;

Query Match 30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 51;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSLSQTLRD1QRM 24
DB 224 GLRTEGLFRASASQIVREIQRL 246

RESULT 10
Q8TGZ1
ID Q8TGZ1 PRELIMINARY; PRT; 818 AA.
AC Q8TGZ1
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE Archaea-specific superfamily II helicase.
GN MK0835.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Nataré D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RA "The complete genome of Hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
DR EMBL: AE010374; AAM02048.1; -.
DR InterPro; IPR003593; AAA_Attrase.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICG; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 818 AA; 91715 MW; C2136200A710817E CRC64;

Query Match 30.9%; Score 55; DB 17; Length 818;
Best Local Similarity 54.2%; Pred. No. 98;
Matches 13; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSLSQTLRD1QRM 25
DB 2 GLRTEGLFRASASQIVREIQRL 246

```

---

```

Db 134 GFSQTLLEKLERLLHELDIDRVV 157

RESULT 11
Q8VD56
ID Q8VD56 PRELIMINARY; PRT; 155 AA.
AC Q8VD56
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Putative bHLH transcription factor (Fragment).
GN SGN-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley;
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S., Ohno K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.;
RT "Sgn1, a Basic Helix-Loop-Helix Transcription Factor Delineates the
RL Salivary Gland Duct Cell Lineage in Mice."
RL Dev. Biol. 240:517-530 (2001).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL: AB046449; BAB83912.1; -.
DR InterPro; IPR001092; HLH_Basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00039; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
FT NON TER 1
SQ SEQUENCE 155 AA; 18146 MW; 6EBADD1A16B88D6 CRC64;

Query Match 30.1%; Score 53.5; DB 11; Length 155;
Best Local Similarity 36.1%; Pred. No. 28;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;

QY 3 LSOEQLEHRSLSQTLRD-----IQRLFPDEKE 31
DB 100 LPEDYLEKLSKVETLRAAKIYSYQSLYPDETE 135

RESULT 12
Q14687
ID Q14687 PRELIMINARY; PRT; 4957 AA.
AC Q14687
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Cnaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RL trithorax."
RL Oncogene 15:549-560 (1997).
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL: AF010404; AAC51735.1; -.
DR InterPro; IPR003889; FYRICH_C.
DR InterPro; IPR003888; FYRICH_N.
DR InterPro; IPR000910; HMG_12_box.

```

2

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities";  
 RL Nature 417:453-463(2002).  
 DR EMBL; AE011625; AAM34916.1; --  
 DR InterPro; IPR002886; Peptidase\_M37.  
 DR Pfam; PF01551; Peptidase\_M37; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 411 AA; 44337 MW; E193496126355426 CRC64;

Query Match 29.8%; Score S3; DB 16; Length 411;  
 Best Local Similarity 36.4%; Pred. NO. 89;  
 Matches 12; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDIQRMFLFPDEKFTG 34  
 DB 17 GASQSQRETERKQLQRLADELKTISADRELEG 49

Search completed: November 13, 2003, 09:27:39  
 Job time : 14.4397 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 11.4397 Seconds  
(without alignments)  
388.502 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136

Sequence: 1 VVVFSTENAKAAEAVLKQGVETIVSPH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_19Jun03.\*
- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
  - 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
  - 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	1426	23	AAB711229
2	136	100.0	1435	22	AB811808
3	129	94.9	1440	22	AA005955
4	98	72.1	1429	22	AB858779
5	98	72.1	1464	23	AB871228
6	94	69.1	1494	23	AAU78460
7	91	66.9	114	23	ABP06595
8	67	49.3	320	23	AAU78461
9	65	47.8	113	22	AA007544

10	52	38.2	360	21	AA833446	Zea mays protein f
11	52	38.2	448	21	AA833445	Zea mays protein f
12	52	38.2	509	21	AA833444	Zea mays protein f
13	51	37.5	1049	22	AB860387	Drosophila melanog
14	50	36.8	130	23	AB889793	Human polypeptide
15	50	36.8	365	22	AB815088	Novel human diago
16	50	36.8	621	23	AAE15740	Human aminoacyl-tr
17	50	36.8	631	21	AA843285	Human ORFX ORF1049
18	50	36.8	1063	23	AB808919	Human aminoacyl tr
19	49	36.0	187	23	ABP66271	Blidobacterium lo
20	49	36.0	984	24	ABU25689	Aspergillus fumiga
21	49	36.0	1058	24	ABJ26489	Aspergillus fumiga
22	48	35.3	498	21	AA817973	Arabidopsis thalia
23	48	35.3	498	21	AA849364	Arabidopsis thalia
24	48	35.3	609	21	AA817972	Arabidopsis thalia
25	48	35.3	609	21	AA849363	Arabidopsis thalia
26	48	35.3	630	21	AA817971	Arabidopsis thalia
27	48	35.3	630	21	AA849362	Arabidopsis thalia
28	47	34.6	886	21	AA840431	Arabidopsis thalia
29	47	34.6	905	21	AA839800	Arabidopsis thalia
30	47	34.6	1095	21	AA840430	Arabidopsis thalia
31	47	34.6	1114	21	AA839799	Arabidopsis thalia
32	47	34.6	1168	21	AA840429	Arabidopsis thalia
33	47	34.6	1187	21	AA839798	Arabidopsis thalia
34	47	34.6	2639	20	AAW73476	Grapevine leafroll
35	45.5	33.5	341	22	ABG70889	Drosophila melanog
36	45	33.1	401	22	ABG04882	Novel human diago
37	44.5	32.7	119	21	AA808467	Arabidopsis thalia
38	44.5	32.7	137	21	AA851008	Arabidopsis thalia
39	44.5	32.7	152	21	AA808466	Arabidopsis thalia
40	44.5	32.7	160	21	AA808465	Arabidopsis thalia
41	44.5	32.7	170	21	AA851007	Arabidopsis thalia
42	44.5	32.7	170	21	AA877970	A. thaliana enviro
43	44.5	32.7	177	21	AA851006	Arabidopsis thalia
44	44.5	32.7	447	22	AA882903	S. epidermidis ope
45	44.5	32.7	450	22	AA883066	S. epidermidis ope

## ALIGNMENTS

### RESULT 1

AA871229	18-NOV-2002 standard; Protein; 1426 AA.
ID	AA871229 standard; Protein; 1426 AA.
XX	AA871229;
AC	AA871229;
DT	18-NOV-2002 (first entry)
XX	Human legless homologue lgs/bcl9 protein.
DE	Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW	tissue proliferation; tumour; cytosolic; cellular disorder; colon;
KW	blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW	medulloblastoma; skin cancer; tissue regeneration; tissue repair.
OS	Homo sapiens.
XX	XX
PN	US2002086986-A1.
XX	XX
PD	04-JUL-2002.
XX	XX
PF	27-JUL-2001; 2001US-0915543.
XX	XX
PR	28-JUL-2000; 2000US-221502P.
XX	XX
PA	(BASL/) BASLER K.
PA	(BRUN/) BRUNNER E.
PA	(FROB/) FROESCH B.
PA	(KRAM/) KRAMPS T.
PA	(PETE/) PETER O.
XX	XX
PI	Basler K, Brunner E, Froesch B, Kramps T, Peter O;

XX WPI; 2002-635689/68.  
 DR N-PSDB; AAF98467.  
 XX Novel polypeptide useful in therapeutic method for treating disorders  
 PT of cell fate such as cell differentiation or cell proliferation -  
 XX Example II; Fig 8B; 41pp; English.  
 XX This invention describes a novel polypeptide sharing one or more  
 CC homologous amino acid domains with the legless (lgs) protein, a  
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
 CC involved in the formation and maintenance of spatial arrangements  
 CC and proliferation of tissues during development, and in the formation  
 CC and growth of many human tumours. The products of the invention have  
 CC cytoskeletal activity and can be used to treat cellular disorders, blood  
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,  
 CC where the cancerous condition is colon, breast, head and neck, brain,  
 CC thyroid, medulloblastoma or skin cancer. The product could also be used  
 CC to promote tissue regeneration and repair. This sequence represents the  
 CC human legless (lgs) protein homologue lgs/bcl9 described in the  
 CC disclosure of the invention.  
 XX SQ Sequence 1426 AA;  
 Query Match 100.0%; Score 136; DB 23; Length 1426;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28  
 DB 177 VYVFSTEMANKAAEAVLKQGVETIVSFH 204  
 RESULT 2  
 ABBI1808  
 ID ABB11808 standard; peptide; 1435 AA.  
 AC ABB11808;  
 XX 11-JAN-2002 (first entry)  
 XX Human BCL9 homologue, SEQ ID NO:2178.  
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; anti-inflammation;  
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnerary; antiulcer.  
 XX Homo sapiens.  
 OS  
 XX W0200157188-A2.  
 XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US03800.  
 XX 03-FEB-2000; 2000US-0496914.  
 XX 27-APR-2000; 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Drmanac RT;  
 XX

DR WPI; 2001-457740/49.  
 DR N-PSDB; ABA09052.  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX Claim 20; Page 256-257; 1963pp; English.  
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX SQ Sequence 1435 AA;  
 Query Match 100.0%; Score 136; DB 22; Length 1435;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28  
 DB 217 VYVFSTEMANKAAEAVLKQGVETIVSFH 244  
 RESULT 3  
 AAO05855  
 ID AAO05855 standard; Protein; 140 AA.  
 XX AAO05855;  
 XX 06-NOV-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 19747.  
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX



The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in APP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for cut

CC Protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 114 AA;

Query Match 66.9%; Score 91; DB 23; Length 114;  
Best Local Similarity 57.1%; Pred. No. 1.7e-06;  
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28

DB 78 VYVFTHLANTAEEAVLQGRADSILAYH 105

RESULT 8

AAU78461

ID AAU78461 standard; Protein; 320 AA.

XX

AC AAU78461;

XX

DT 02-JUL-2002 (first entry)

XX

DE Mouse beta-catenin nuclear localised protein #2.

XX

KW Mouse; beta-catenin nuclear localised protein; cancer;

KW gene therapy; EST; expressed sequence tag.

XX

OS Mus musculus.

XX

PN WO200224738-A1.

XX

PD 28-MAR-2002.

XX

PF 19-SEP-2001; 2001WO-JP08140.

XX

PR 22-SEP-2000; 2000JP-0287876.

XX

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX

PI Akiyama T, Adachi S;

XX

DR WP1; 2002-330014/36.

XX

DR N-PSDB; ABK47632.

XX

PT New beta-catenin nuclear localised protein for diagnosis and treatment

PT of diseases associated with nuclear localisation of beta-catenin e.g.

PT cancer -

XX

PS Claim 2; Page 91-92; 113pp; Japanese.

XX

CC The invention relates to a beta-catenin nuclear localised protein  
CC and DNA encoding the protein. The protein and encoding DNA are  
CC applicable in diagnosis and treatment of diseases associated with  
CC nuclear localisation of beta-catenin e.g. cancer, including gene  
CC therapy. The present sequence represents the amino acid sequence of  
CC mouse beta-catenin nuclear localised protein #2.

XX

SQ Sequence 320 AA;

Query Match 49.3%; Score 67; DB 23; Length 320;

Best Local Similarity 57.1%; Pred. No. 0.035;

Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKQGVETIVSFH 28

DB 1 LANTAEEAVLQGRADSILAYH 21

RESULT 9

AAO07544  
ID AAO07544 standard; Protein; 113 AA.

XX

AC AAO07544;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 21436.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US04927.

XX

PR 28-FEB-2000; 2000US-0515126.

XX

PR 18-MAY-2000; 2000US-0577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WP1; 2001-514838/56.

XX

DR N-PSDB; AAI87475.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX

PS Claim 20; SEQ ID NO 21436; 1399pp + Sequence Listing; English.

XX

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 113 AA;

Query Match 47.8%; Score 65; DB 22; Length 113;

Best Local Similarity 75.0%; Pred. No. 0.021;

Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKQGVETIVSF 27

DB 81 MCVOAAEAVLKQGVETDSF 100

RESULT 10

AAG33446

ID AAG33446 standard; Protein; 360 AA.

XX

AC AAG33446;

XX

DT 18-OCT-2000 (first entry)

XX

DE 2ea mays protein fragment SEQ ID NO: 40525.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX Zea mays subsp. mays.  
XX EP1033405-A2.  
XX PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130810.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131149.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135529.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143544.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

```
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161320.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 52; DB 21; Length 360;
Best Local Similarity 50.0%; Pred. No. 9.4;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSP 27
Db 97 FSTRLLNNLENLVLRGPETIAAF 120

RESULT 11
AAG33445
ID AAG33445 standard; Protein; 448 AA.
XX
AC AAG33445;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 40524.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
```

```
PR 06-SEP-2000.
PR 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
```

XX  
PR 25-FEB-1999; 99US-0121825.



PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139494.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 52; DB 21; Length 509;

Best Local Similarity 50.0%; Pred. No. 14;

Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQVETIVSF 27

DB 246 FSTRLLNNLENLVLRGSPETIAAF 269

RESULT 13

ABB60387

ID ABB60387 standard; Protein; 1049 AA.

XX AC ABB60387;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 7953.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX DR N-PSDB; ABL04490.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Disclosure; SEQ ID NO 7953; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1049 AA;

Query Match 37.5%; Score 51; DB 22; Length 1049;

Best Local Similarity 40.7%; Pred. No. 49;

Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVSTEMANKAAEAVLKQVETIVSFH 28

DB 471 YVSCDMAASATEAVRSGELKIIPHH 497

RESULT 14

ABB89793

ID ABB89793 standard; Protein; 130 AA.

XX AC ABB89793;

XX DT 24-MAY-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 2169.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder; neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US16450.

XX PR 19-MAY-2000; 2000US-205515P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

XX DR N-PSDB; ABL90202.

XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders -

XX PS Claim 11; SEQ ID NO 2169; 2081pp + Sequence Listing; English.

XX CC The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (ii) and its binding partners are useful in medical imaging of sites expressing (i). (i) and (ii) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AIG300010-AIG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct/sequences](http://ftp.wipo.int/pub/published/pct/sequences).

```
Query MatCh      36.8%; Score 50; DB 22; Length 365;
Best Local Similarity 29.6%; Pred. No. 20;
Matches 8; Conservative 8; Mismatches 11; Indels 0; Gaps 0;
```

**Qy** 2 YVFSTEMANKAAEAVLKQVETIVSPH 28  
||| : | : : : | : : |  
**Dd** 67 YVFAYPVCNASKATIISGLTECLIHCH 93

Search completed: November 13, 2003, 09:25:19  
Job time : 12.4397 secs

Real5088  
IID ABG15088 standard; Protein; 365 AA.  
XX  
XX  
ABG15088;  
XX  
XX AAC  
18-FEB-2002 (first entry)  
DT  
XX  
XX  
DE Novel human diagnostic protein #15079.  
XX  
XX  
Human; chromosome mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

ABG15088;  
18-FEB-2002 (first entry)

18-FEB-2002 (first entry)  
Novel human diagnostic protein #15079.

XX	Novel human diagnostic protein #15079.
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
DE	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

AA	Human; chromo
KW	food suppleme
KW	
XX	Homo sapiens.
OS	

OS  
XX  
PN

WO200175067-A2.  
11-OCT-2001.

PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.

30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.

31-MAR-2000; 2000US-0540217.  
23-AUG-2000; 2000US-0649167.  
XX

XX  
PA (HYSE-) HYSEQ INC.  
XX

XX Drmanac RT, Liu C, Tang YT  
PI  
XX

XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS79275.

XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS79275.  
YY

DR N-PSDB; AAS79275.  
XX  
PT New isolated polynucleotide a

New isolated polynucleotide and encoded polypeptides, useful in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits and to assess  
 biodiversity -  
 Claim 20; SEQ ID NO 45447; 103pp; English.

Claim 20; SEQ ID No 45447; 103pp; English.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:19 ; Search time 4.14008 Seconds  
(without alignments)  
286.155 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136  
Sequence: 1 VYVFTEWANKAAEVLKQVETIVSFH 28

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCUTS COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	47	34.6	2639	3 US-09-080-983-3	Sequence 3, Appli
2	44	32.4	718	3 US-09-090-808-2	Sequence 2, Appli
3	44	32.4	718	3 US-09-447-453-2	Sequence 2, Appli
4	43	31.6	145	4 US-09-091-725-37	Sequence 37, Appl
5	43	31.6	438	3 US-08-486-099-93	Sequence 93, Appl
6	43	31.6	438	3 US-08-360-107A-103	Sequence 103, App
7	43	31.6	438	3 US-08-484-223B-93	Sequence 93, Appl
8	43	31.6	438	3 US-08-919-597-93	Sequence 93, Appl
9	43	31.6	438	3 US-08-475-668A-93	Sequence 93, Appl
10	43	31.6	438	3 US-08-485-551A-93	Sequence 93, Appl
11	43	31.6	438	3 US-08-471-913A-93	Sequence 93, Appl
12	43	31.6	438	3 US-08-485-264A-93	Sequence 93, Appl
13	43	31.6	438	4 US-08-474-349A-93	Sequence 93, Appl
14	43	31.6	438	4 US-08-255-208A-29	Sequence 29, Appl
15	43	31.6	438	4 US-08-470-896-93	Sequence 93, Appl
16	43	31.6	438	4 US-08-485-546A-93	Sequence 93, Appl
17	43	31.6	662	1 US-08-224-657-88	Sequence 88, Appl
18	43	31.6	662	4 US-09-354-138-88	Sequence 88, Appl
19	42	30.9	154	4 US-08-732-210-103	Sequence 103, App
20	42	30.9	154	4 US-09-732-210-210	Sequence 210, App
21	42	30.9	188	4 US-09-252-991A-29737	Sequence 29737, A
22	42	30.9	300	4 US-08-634-238-219	Sequence 219, App
23	42	30.9	358	4 US-09-934-901-8	Sequence 8, Appli
24	42	30.9	541	4 US-09-627-376-11	Sequence 11, Appl
25	42	30.9	1820	3 US-07-998-289B-8	Sequence 8, Appli
26	41.5	30.5	111	4 US-08-775-932-20	Sequence 20, Appl
27	41.5	30.5	678	1 US-08-844-085-2	Sequence 2, Appli

28	41.5	30.5	745	1 US-08-453-472-5	Sequence 5, Appli
29	41.5	30.5	745	1 US-08-038-948-9	Sequence 9, Appli
30	41.5	30.5	745	1 US-08-453-952-5	Sequence 5, Appli
31	41.5	30.5	745	2 US-08-484-993B-43	Sequence 43, Appli
32	41.5	30.5	745	2 US-08-862-903-5	Sequence 5, Appli
33	41.5	30.5	745	2 US-08-484-158B-43	Sequence 43, Appli
34	41.5	30.5	745	2 US-08-484-596A-43	Sequence 43, Appli
35	41.5	30.5	745	2 US-08-480-150A-43	Sequence 43, Appli
36	41.5	30.5	745	3 US-08-458-731-43	Sequence 43, Appli
37	41.5	30.5	745	3 US-08-149-223A-43	Sequence 43, Appli
38	41	30.1	111	4 US-09-107-532A-5449	Sequence 5449, Ap
39	41	30.1	162	4 US-09-134-001C-3670	Sequence 3670, Ap
40	41	30.1	211	3 US-09-075-194-1	Sequence 1, Appli
41	41	30.1	256	2 US-08-578-516-6	Sequence 6, Appli
42	41	30.1	378	4 US-09-107-532A-6500	Sequence 6500, Ap
43	41	30.1	569	4 US-09-107-532A-6689	Sequence 6689, Ap
44	41	30.1	676	4 US-09-107-532A-5806	Sequence 5806, Ap
45	41	30.1	689	4 US-09-499-964-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-080-983-3  
; Sequence 3, Application US/09080983  
; Patent No. 6197948  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Hai-Ying  
; APPLICANT: Ling, Kai-Shu  
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS TYPE 2 PROTEINS  
; TITLE OF INVENTION: AND THEIR USES  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devane & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/080,983  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/047,194  
; FILING DATE: 20-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1631  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2639 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-080-983-3

Query Match 34.6%; Score 47; DB 3; Length 2639;  
Best Local Similarity 33.3%; Pred. No. 1.3e+02;  
Matches 9; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 95203620.0  
 FILING DATE: 22-DEC-1995  
 APPLICATION NUMBER: EP 96200943.7  
 FILING DATE: 11-APR-1996  
 ATTORNEY/AGENT INFORMATION:

NAME: E. Victor Donahue  
REGISTRATION NUMBER: 35,492  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-091-725-37

Query Match 31.6%; Score 43; DB 4; Length 145;  
Best Local Similarity 43.8%; Pred. No. 16;  
Matches 14; Conservative 4; Mismatches 6; Indels 8; Gaps 2;

Qy 5 STEMANKAA-BAVLKG-----QVETVVSFH 28  
Db 15 SADVKAKAAKALNGTOSTRKVRTSVSFH 46

RESULT 5  
US-08-486-099-93  
; Sequence 93, Application US/08486099  
; Patent No. 6013283  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS  
; TITLE OF INVENTION: B VIRUS TRANSMISSION  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,099  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-031  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 93:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-486-099-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAA-BAVLKGQVETIVS 26

Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 6  
US-08-360-107A-103  
; Sequence 103, Application US/08360107A  
; Patent No. 6017536  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
; TITLE OF INVENTION: TRANSMISSION  
; NUMBER OF SEQUENCES: 149  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/360,107A  
; FILING DATE: 20-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 103:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-360-107A-103

Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAA-BAVLKGQVETIVS 26  
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 7  
US-08-484-223B-93  
; Sequence 93, Application US/08484223B  
; Patent No. 6020459  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.

```

; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-223B-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAAEAVLKQGVETIVS 26
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 8
US-08-919-597-93
; Sequence 93, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-919-597-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAAEAVLKQGVETIVS 26
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 9
US-08-475-668A-93
; Sequence 93, Application US/08475668A
; Patent No. 6060065
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,668A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 438 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-08-475-668A-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
 Best Local Similarity 36.4%; Pred. No. 63;  
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAAEAVLKQGVETIVS 26  
 ||||| : : : : :  
 Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 10  
 US-08-485-551A-93  
 ; Sequence 93, Application US/0848551A  
 ; Patent No. 6068973

GENERAL INFORMATION:  
 APPLICANT: Bolognesi, Dani P.  
 APPLICANT: Matthews, Thomas J.  
 APPLICANT: Wild, Carl T.  
 APPLICANT: Barney, Shawn O.  
 APPLICANT: Lambert, Dennis M.  
 APPLICANT: Petteway, Stephen R.  
 APPLICANT: Langlois, Alphonse J.  
 TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE  
 TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
 TITLE OF INVENTION: TRANSMISSION  
 NUMBER OF SEQUENCES: 211  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/485,551A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 7872-023  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-9741/8664  
 TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 93:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 438 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-08-485-551A-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
 Best Local Similarity 36.4%; Pred. No. 63;  
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAAEAVLKQGVETIVS 26  
 ||||| : : : : :  
 Db 41 SLEQSNKAIEIREATQETVIA 62

## RESULT 11

US-08-471-913A-93  
 ; Sequence 93, Application US/08471913A  
 ; Patent No. 6093794

## GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.  
 APPLICANT: Matthews, Thomas J.  
 APPLICANT: Wild, Carl T.  
 APPLICANT: Barney, Shawn O.  
 APPLICANT: Lambert, Dennis M.  
 APPLICANT: Petteway, Stephen R.  
 APPLICANT: Langlois, Alphonse J.

TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
 TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS  
 TITLE OF INVENTION: TRANSMISSION  
 NUMBER OF SEQUENCES: 214

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/471,913A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 7872-030  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 93:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 438 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-08-471-913A-93

Query Match 31.6%; Score 43; DB 3; Length 438;

Best Local Similarity 36.4%; Pred. No. 63;  
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 5 STEMANKAAEAVLKQGVETIVS 26

||||| : : : : :  
 Db 41 SLEQSNKAIEIREATQETVIA 62

## RESULT 12

US-08-485-264A-93  
 ; Sequence 93, Application US/08485264A  
 ; Patent No. 6228983

## GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.  
 APPLICANT: Matthews, Thomas J.  
 APPLICANT: Wild, Carl T.  
 APPLICANT: Barney, Shawn O.  
 APPLICANT: Lambert, Dennis M.  
 APPLICANT: Petteway, Stephen R.  
 APPLICANT: Langlois, Alphonse J.

TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
 TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING



;; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION  
;; NUMBER OF SEQUENCES: 232  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds LLP  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/485,264A  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 7872-021  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 93:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 438 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; US-08-485-264A-93  
;;  
Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
QY 5 STEMANKAAEAVLKGQVETIVS 26  
DB 41 SLEQSNKAIEIREATQETVIA 62  
;;  
RESULT 13  
US-08-474-349A-93  
;; Sequence 93, Application US/08/74349A  
;; Patent No. 6333395  
;; GENERAL INFORMATION:  
;; APPLICANT: Bolognesi, Dani P.  
;; APPLICANT: Matthews, Thomas J.  
;; APPLICANT: Wild, Carl T.  
;; APPLICANT: Barney, Shawn O.  
;; APPLICANT: Lambert, Dennis M.  
;; APPLICANT: Petteway, Stephen R.  
;; APPLICANT: Langlois, Alphonse J.  
;; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
;; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA  
;; TITLE OF INVENTION: VIRUS TRANSMISSION  
;; NUMBER OF SEQUENCES: 517  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/474,349A  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 7872-024  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 93:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 438 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; US-08-474-349A-93  
;;  
Query Match 31.6%; Score 43; DB 4; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
QY 5 STEMANKAAEAVLKGQVETIVS 26  
DB 41 SLEQSNKAIEIREATQETVIA 62  
;;  
RESULT 14  
US-08-255-208A-29  
;; Sequence 29, Application US/08255208A  
;; Patent No. 6440656  
;; GENERAL INFORMATION:  
;; APPLICANT: Bolognesi, Dani P.  
;; APPLICANT: Matthews, Thomas J.  
;; APPLICANT: Wild, Carl T.  
;; APPLICANT: Barney, Shawn O.  
;; APPLICANT: Lambert, Dennis M.  
;; APPLICANT: Petteway Jr., Stephen R.  
;; TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV  
;; TITLE OF INVENTION: TRANSMISSION  
;; NUMBER OF SEQUENCES: 111  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds LLP  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10036  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/255,208A  
;; FILING DATE: 07-JUN-1994  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 7872-010  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 29:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 438 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein

US-08-255-208A-29

Query Match 31.6%; Score 43; DB 4; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQVETIVS 26  
DB 41 SLEQSNKAIEIREATQETVIA 62

RESULT 15

US-08-470-896-93  
Sequence 93, Application US/08470896  
Patent No. 6479055

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Pettaway, Stephen R.  
APPLICANT: Langlois, Alphonse J.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,896  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:  
LENGTH: 438 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

US-08-470-896-93

Query Match 31.6%; Score 43; DB 4; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQVETIVS 26  
DB 41 SLEQSNKAIEIREATQETVIA 62

Search completed: November 13, 2003, 09:28:23  
Job time: 5.14008 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:27:44 ; Search time 7.51751 Seconds  
(without alignments)  
679.968 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136  
Sequence: 1 VYVSTEMANKAAEAVLKQVETIVSFH 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues  
Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	136	100.0	28	15	US-10-322-579-3
2	136	100.0	1426	15	US-10-322-579-15
3	98	72.1	28	15	US-10-322-579-2
4	49	36.0	984	15	US-10-128-714-3547
5	49	36.0	1058	15	US-10-128-714-8547
6	47	34.6	732	14	US-10-008-355-6
7	46	33.8	471	12	US-10-167-547C-8
8	45	33.1	507	12	US-10-167-547C-10
9	45	33.1	747	15	US-10-156-761-10721
10	44.5	32.7	170	12	US-10-342-224-92
11	44	32.4	171	15	US-10-156-761-9002
12	44	32.4	484	9	US-09-815-242-10457
13	44	32.4	718	14	US-10-117-846-2
14	43	31.6	80	11	US-09-764-891-5144
15	43	31.6	89	12	US-10-029-386-28376

16	43	31.6	134	12	US-10-029-386-31669
17	43	31.6	315	12	US-10-306-762-212
18	43	31.6	602	12	US-10-250-381-7
19	43	31.6	623	12	US-10-032-585-7819
20	43	31.6	662	11	US-09-951-061A-141
21	42	30.9	248	12	US-10-097-111-304
22	42	30.9	277	15	US-10-156-761-8036
23	42	30.9	326	15	US-10-156-761-9336
24	42	30.9	358	10	US-09-934-501-8
25	42	30.9	358	10	US-09-934-868-18
26	42	30.9	358	12	US-10-320-874-8
27	42	30.9	358	16	US-10-320-924-8
28	42	30.9	380	12	US-10-183-708-42
29	42	30.9	380	12	US-09-932-227-42
30	42	30.9	399	12	US-10-293-971-9
31	42	30.9	541	14	US-10-047-876A-11
32	41.5	30.5	111	10	US-09-775-932-20
33	41.5	30.5	469	8	US-08-781-986A-5245
34	41	30.1	227	9	US-09-815-242-11197
35	41	30.1	279	9	US-09-861-451A-28
36	41	30.1	370	10	US-09-712-363-187
37	41	30.1	370	12	US-10-361-460-10
38	41	30.1	451	11	US-09-910-186A-12
39	41	30.1	456	9	US-09-815-242-4391
40	41	30.1	456	9	US-09-815-242-10636
41	41	30.1	516	10	US-09-738-626-5086
42	41	30.1	568	15	US-10-156-761-11478
43	41	30.1	638	15	US-10-234-432-59
44	41	30.1	689	14	US-10-052-664-1
45	41	30.1	689	15	US-10-097-340-221

ALIGNMENTS

RESULT 1

US-10-322-579-3  
; Sequence 3, Application US/10322579  
; Publication No. US20030114413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMES, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Human lgs/bcl9  
US-10-322-579-3

Query Match 100.0%; Score 136; DB 15; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1e-14; Indels 0; Gaps 0;  
Matches 28; Conservative 0; Mismatches 0;  
Qy 1 VYVSTEMANKAAEAVLKQVETIVSFH 28  
Db 1 VYVSTEMANKAAEAVLKQVETIVSFH 28

RESULT 2

US-10-322-579-15

Sequence 15, Application US/10322579  
Publication No. US20030114413A1  
GENERAL INFORMATION:  
APPLICANT: BASLER, Konrad  
APPLICANT: BRUNNER, Erich  
APPLICANT: FROESCH, Barbara  
APPLICANT: KRAMPS, Thomas  
APPLICANT: PETER, Oliver  
TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
FILE REFERENCE: Q60361  
CURRENT APPLICATION NUMBER: US/10/322,579  
CURRENT FILING DATE: 2002-12-19  
PRIOR APPLICATION NUMBER: US/09/915,543  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: 60/221,502  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 15  
LENGTH: 1426  
TYPE: PRT  
ORGANISM: Human lgs/bc19  
US-10-322-579-15

Query Match 100.0%; Score 136; DB 15; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 1.1e-12;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYFSTEMANKAAEAVLKQVETIVSPH 28  
DB 177 VYFSTEMANKAAEAVLKQVETIVSPH 204

## RESULT 3

US-10-322-579-2  
Sequence 2, Application US/10322579  
Publication No. US20030114413A1  
GENERAL INFORMATION:  
APPLICANT: BASLER, Konrad  
APPLICANT: BRUNNER, Erich  
APPLICANT: FROESCH, Barbara  
APPLICANT: KRAMPS, Thomas  
APPLICANT: PETER, Oliver  
TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
FILE REFERENCE: Q60361  
CURRENT APPLICATION NUMBER: US/10/322,579  
CURRENT FILING DATE: 2002-12-19  
PRIOR APPLICATION NUMBER: US/09/915,543  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: 60/221,502  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 28  
TYPE: PRT  
ORGANISM: Drosophila lgs  
US-10-322-579-2

Query Match 72.1%; Score 98; DB 15; Length 28;  
Best Local Similarity 57.1%; Pred. No. 1.1e-08;  
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYFSTEMANKAAEAVLKQVETIVSPH 28  
DB 1 IFVSTQLANKAESVLGQFQTTIAYH 28

## RESULT 4

US-10-128-714-3547  
Sequence 3547, Application US/10128714

Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroshkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
TITLE OF INVENTION: Methods of Use  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3547  
LENGTH: 984  
TYPE: PRT  
ORGANISM: Aspergillus fumigatus  
US-10-128-714-3547

Query Match 36.0%; Score 49; DB 15; Length 984;  
Best Local Similarity 52.6%; Pred. No. 52;  
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKQV 21  
DB 500 VYRANKAAEAVLKSKL 518

## RESULT 5

US-10-128-714-8547  
Sequence 8547, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroshkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
TITLE OF INVENTION: Methods of Use  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8547  
LENGTH: 1058  
TYPE: PRT  
ORGANISM: Aspergillus fumigatus  
US-10-128-714-8547

Query Match 36.0%; Score 49; DB 15; Length 1058;  
Best Local Similarity 52.8%; Pred. No. 57;  
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 VFSTEMANKAAEAVLKGVQV 21  
| : ||| : ||| : ||| : ||| :  
Db 500 VYFANWANKSRAAVLKSKL 518

## RESULT 6

US-10-008-355-6  
; Sequence 6, Application US/10008355  
; Publication No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 6  
; LENGTH: 732  
; TYPE: PRT  
; ORGANISM: Shewanella putrefaciens  
US-10-008-355-6

Query Match 34.6%; Score 47; DB 14; Length 732;  
Best Local Similarity 39.1%; Pred. No. 76;  
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGVET 23  
| : : : : : : : : : : : : : : : : : :  
Db 517 VALYDTNMAOEKAEKILAGKLS 539

## RESULT 7

US-10-167-547C-8  
; Sequence 8, Application US/10167547C  
; Publication No. US20030170653A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. du Pont de Nemours and Company  
; APPLICANT: Damude, Howard G.  
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gamma  
; TITLE OF INVENTION: Butyrolactone and its Intermediates  
; FILE REFERENCE: CL1804 US NA  
; CURRENT APPLICATION NUMBER: US/10/167,547C  
; CURRENT FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: 60/297198  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: Microsoft Office 07  
; SEQ ID NO 8  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: tulip pistil  
US-10-167-547C-8

Query Match 33.8%; Score 46; DB 12; Length 471;  
Best Local Similarity 41.7%; Pred. No. 64;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 4 FSTEMANKAAEAVLKGVETVSF 27  
| : : : : : : : : : : : : : : : : : :  
Db 204 FSTRLANNLEKILITEGPETIAAF 227

## RESULT 8

; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress

US-10-167-547C-10  
; Sequence 10, Application US/10167547C  
; Publication No. US20030170653A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. du Pont de Nemours and Company  
; APPLICANT: Damude, Howard G.  
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gamma  
; TITLE OF INVENTION: Butyrolactone and its Intermediates  
; FILE REFERENCE: CL1804 US NA  
; CURRENT APPLICATION NUMBER: US/10/167,547C  
; CURRENT FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: 60/297198  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: Microsoft Office 07  
; SEQ ID NO 10  
; LENGTH: 507  
; TYPE: PRT  
; ORGANISM: Alstroemeria  
US-10-167-547C-10

Query Match 33.1%; Score 45; DB 12; Length 507;  
Best Local Similarity 37.5%; Pred. No. 1e-02;  
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 4 FSTEMANKAAEAVLKGVETVSF 27  
| : : : : : : : : : : : : : : : : : :  
Db 246 FSTRLANNLEKILITEGPETIAAF 269

## RESULT 9

US-10-156-761-10721  
; Sequence 10721, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10721  
; LENGTH: 747  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10721

Query Match 33.1%; Score 45; DB 15; Length 747;  
Best Local Similarity 39.1%; Pred. No. 1.6e+02;  
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 3 VFSTEMANKAAEAVLKGVETIV 25  
| : : : : : : : : : : : : : : : : : :  
Db 250 VFATSSLNRRARQALDRGETERAV 272

## RESULT 10

US-10-342-224-92  
; Sequence 92, Application US/10342224  
; Publication No. US20030162294A1  
; GENERAL INFORMATION:  
; APPLICANT: Nathalie Verbruggen  
; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress

FILE REFERENCE: CNN-012US  
CURRENT APPLICATION NUMBER: US/10/342,224  
CURRENT FILING DATE: 2003-01-13  
PRIOR APPLICATION NUMBER: US/09/762,154  
PRIOR FILING DATE: 2002-02-02  
PRIOR APPLICATION NUMBER: EP 98202634.6  
PRIOR FILING DATE: 1998-08-04  
NUMBER OF SEQ ID NOS: 123  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 92  
LENGTH: 170  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-342-224-92

Query Match 32.4%; Score 44.5; DB 12; Length 170;  
Best Local Similarity 52.0%; Pred. No. 33;  
Matches 13; Conservative 0; Mismatches 7; Indels 5; Gaps 1;  
QY 6 TEMANKAAEAVL-----KGQVETIV 25  
Db 85 TEYVKTVEVLTDLTKKQVETIV 109

RESULT 11  
US-10-156-761-9002  
Sequence 9002, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIRA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 9002  
LENGTH: 171  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-9002

Query Match 32.4%; Score 44; DB 15; Length 171;  
Best Local Similarity 35.7%; Pred. No. 40;  
Matches 10; Conservative 5; Mismatches 13; Indels 0; Gaps 0;  
QY 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28  
Db 29 IFCYRSMMAKAEQVLDVQWRGILAVH 56

RESULT 12  
US-09-815-242-10457  
Sequence 10457, Application US/09815242  
Patent No. US2002061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10457  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-815-242-10457

Query Match 32.4%; Score 44; DB 9; Length 484;  
Best Local Similarity 30.0%; Pred. No. 14e+02;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
QY 8 MANKAAEAVLKQVETIVSF 27  
Db 52 LLKRSVSVIIGTKTIIGF 71

RESULT 13  
US-10-117-846-2  
Sequence 2, Application US/10117846  
Publication No. US20020168673A1  
GENERAL INFORMATION:  
APPLICANT: Fuller, Margaret T  
APPLICANT: Hales, Karen G.  
APPLICANT: Santel, Ansgar H.  
TITLE OF INVENTION: Mitofusins, Fzo Homologs and Functional  
Derivatives Thereof  
FILE REFERENCE: STAN-063CIP3  
CURRENT APPLICATION NUMBER: US/10/117,846  
CURRENT FILING DATE: 2002-04-05  
PRIOR APPLICATION NUMBER: 09/413,285  
PRIOR FILING DATE: 1999-10-06  
PRIOR APPLICATION NUMBER: PCT/US00/27871  
PRIOR FILING DATE: 2000-10-06  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 718  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-117-846-2

Query Match 32.4%; Score 44; DB 14; Length 718;  
Best Local Similarity 34.6%; Pred. No. 2.2e+02;  
Matches 9; Conservative 5; Mismatches 10; Indels 2; Gaps 1;  
QY 2 YVFS--TEMANKAAEAVLKQVETIV 25  
Db 403 YCFEELTENTQVRGCVLSDIKTLI 428

RESULT 14  
US-09-764-891-5144  
Sequence 5144, Application US/09764891

Search completed: November 13, 2003, 09:39:45  
Job time : 7.51751 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 3.81323 Seconds  
(without alignments)  
706.153 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204  
Perfect score: 136  
Sequence: 1 VYVFSTEMANKAAEAVLKQVETIVSF 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9618682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	44.1	205	2 I40812	porphobilinogen sy
2	55	40.4	634	2 T27465	hypothetical prote
3	51	37.5	243	2 G97070	Zn-dependent hydro
4	49	36.0	210	2 D86398	protein Fl17L21.2 I
5	49	36.0	458	2 F71315	probable response
6	49	36.0	586	2 D82484	Sgat protein VCA02
7	48.5	35.7	311	2 C81380	probable D-2-hydro
8	48	35.3	330	2 E84074	dihydroxyacetone k
9	47	34.6	319	2 T01822	hypothetical prote
10	47	34.6	352	2 F90179	prolidase (Xaa-Pro
11	47	34.6	662	2 AD0623	probable bacteriop
12	46.5	34.2	586	2 T29695	hypothetical prote
13	46	33.8	330	2 S08500	QUTG protein - Eme
14	46	33.8	363	2 C69962	branched-chain fat
15	46	33.8	461	2 E95987	probable aminotran
16	46	33.8	1289	2 F72308	hypothetical prote
17	46	33.8	2123	2 S55089	probable acetyl-Co
18	45.5	33.5	256	2 S04363	class II histocomp
19	45.5	33.5	268	2 A99261	glutamate CoA-tr
20	45.5	33.5	504	2 S54744	cellulase (EC 3.2.
21	45.5	33.5	505	2 S13962	endoglucanase - Br
22	45	33.1	131	2 T72478	hypothetical prote
23	45	33.1	451	2 E96495	hypothetical prote
24	45	33.1	555	2 H96762	hypothetical prote
25	44.5	32.7	48	2 D90907	hypothetical prote
26	44.5	32.7	50	2 F85710	unknown protein en
27	44	32.4	126	2 T43131	hypothetical prote
28	44	32.4	265	2 T14645	hypothetical prote
29	44	32.4	318	2 C64445	conserved hypothet

30	44	32.4	340	2 T19105	phosphate carrier
31	44	32.4	350	2 A85056	probable transposo
32	44	32.4	355	2 AH3455	acridflavin resista
33	44	32.4	437	2 A72498	probable DNA/panto
34	44	32.4	476	2 AG1051	probable transport
35	44	32.4	484	2 A86116	hypothetical prote
36	44	32.4	484	2 D65230	hypothetical 52.9
37	44	32.4	484	2 A98275	hypothetical prote
38	44	32.4	533	2 T05092	probable 1,2-diac
39	44	32.4	609	2 JCS756	vibriolysin (EC 3.
40	44	32.4	619	2 G72709	probable DNA ligas
41	44	32.4	980	2 T39630	valine-tRNA ligase
42	44	32.4	4427	2 PN0637	polyketide synthas
43	43	31.6	119	2 T18644	hypothetical prote
44	43	31.6	159	2 T40440	6,7-dimethyl-8-rib
45	43	31.6	234	2 AD3133	hypothetical prote

## ALIGNMENTS

### RESULT 1

T40812  
porphobilinogen synthase (EC 4.2.1.24) - Clostridium josui (fragment)  
N/Alternate names: delta-aminolevulinic acid dehydratase  
C/Species: Clostridium josui  
C/Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999  
C/Accession: I40812  
R/Fujino, E.; Fujino, T.; Karita, S.; Sakka, K.; Ohmiya, K.  
J. Bacteriol. 177, 5169-5175, 1995  
A/Title: Cloning and sequencing of some genes responsible for porphyrin biosynthesis  
A/Reference number: AS7344; MUID:95394829; PMID:7665501  
A/Accession: I40812  
A/Status: preliminary; translated from GB/EMBL/DBSJ  
A/Molecule type: DNA  
A/Residues: 1-205 <RES>  
A/Cross-references: GB:D28503; NID:G536874; PIDN:BA05863.1; PID:G556484  
C/Genetics:  
C/Suprafamily: porphobilinogen synthase  
C/Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 44.1%; Score 60; DB 2; Length 205;  
Best Local Similarity 46.2%; Pred. No. 0.23;  
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQVETIVSF 27  
||| : ||| ||| ||| : ||| :  
DB 51 YHFSFDMVGRKAEALKAADVKSVLLF 76

### RESULT 2

T27465  
hypothetical protein Y87G2A.m - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T27465  
R.White, S.  
submitted to the EMBL Data Library, September 1999  
A/Reference number: Z20371  
A/Accession: T27465  
A/Status: preliminary; translated from GB/EMBL/DBSJ  
A/Molecule type: DNA  
A/Residues: 1-634 <WLT>  
A/Cross-references: EMBL:AL110500; NID:el542314; PIDN:CA854487.1; CESP:Y87G2A.m  
A/Experimental source: clone Y87G2A  
C/Genetics:  
A/Gene: CESP:Y87G2A.m  
A/Introns: 74/1; 270/1

Query Match 40.4%; Score 55; DB 2; Length 634;  
Best Local Similarity 44.4%; Pred. No. 4.2; Mismatches 2; Indels 13; Gaps 0;



proposed to establish a regional property tax system - a system of property taxes levied by the regional government and distributed to the local governments in the region.

A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp

A;Reference number: AB1250; MUID:20150912; PMID:10688204

A;Accession: CB1390

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-311 <PAR>

A;Cross-references: GB:AL1139075; GB:AL111168; NID:G6967817; PIDN:CA874209.1; PID:G696784

A;Experimental source: serotype O2, strain NCTC 11168

C;Genetics:

A;Gene: Cj0373

C;Superfamily: phosphoglycerate dehydrogenase

Query Match 35.7%; Score 48.5; DB 2; Length 311;

Best Local Similarity 42.4%; Pred. No. 18;

Matches 14; Conservative 6; Mismatches 8; Indels 5; Gaps 2;

Qy 1 YVFTSTEMANKAAEAV---LKGQVET--IVSPH 28

Db 171 IYVYSTGANKADPVHLEKLLKTDIIISIH 203

RESULT 8

E84074

dihydroxyacetone kinase BH3397 [imported] - *Bacillus halodurans* (strain C-125)

C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C;Accession: E84074

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A;Reference number: AB3650; MUID:20512582; PMID:11058132

A;Accession: E84074

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-330 <STO>

A;Cross-references: GB:AP001518; GB:BA000004; NID:G10175792; PIDN:BA807116.1; GSPDB:GN00

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH3397

Query Match 35.3%; Score 48; DB 2; Length 330;

Best Local Similarity 42.3%; Pred. No. 23;

Matches 11; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 2 YVFTSTEMANKAAEAVLKGQVETIVSF 27

Db 270 YVFMNDVANKLTBEGLNQPKVGSF 295

RESULT 9

T01822

hypothetical protein T27D20.16 - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 10-Dec-1999

C;Accession: T01822

R;Edwards, J.; Wollam, C.; Dubbelde, C.

submitted to the EMBL Data Library, August 1998

A;Description: The sequence of A. thaliana T27D20.

A;Reference number: Z1441

A;Accession: T01822

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-319 <EDW>

A;Cross-references: EMBL:AF076274; NID:G3293583; PID:G3377852

A;Experimental source: cultivar Columbia

C;Genetics:

A;Map position: 4

A;Introns: 49/3; 151/3; 210/3; 269/2

A;Note: T27D20.16

C;Superfamily: Arabidopsis hypothetical protein F7N22.18

Query Match 34.6%; Score 47; DB 2; Length 319;

Best Local Similarity 48.0%; Pred. No. 31;

Matches 12; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 5 STEMANK--AAEAVLKGQVETIVSF 27

Db 243 SIELSQRKLAARAAALIANQAEKITSF 267

RESULT 10

F90179

prolidase (Xaa-Pro dipeptidase) (pepQ) [imported] - *Sulfolobus solfataricus*

C;Species: *Sulfolobus solfataricus*

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001

C;Accession: F90179

R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch

Jonng, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A;Description: *Sulfolobus solfataricus* complete genome.

A;Reference number: A99139

A;Accession: F90179

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-352 <KUR>

A;Cross-references: GB:AE006641; NID:G13813507; PIDN:AAK40693.1; GSPDB:GN00155

C;Genetics:

A;Gene: pepQ

C;Superfamily: X-Pro aminopeptidase

Query Match 34.6%; Score 47; DB 2; Length 352;

Best Local Similarity 43.5%; Pred. No. 35;

Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 2 YVFTSTEMANKAAEAVLKGQVETI 24

Db 232 FVFKNSERAKKYEVVLEAQMEAI 254

RESULT 11

AD0623

probable bacteriophage protein STV1061 [imported] - *Salmonella enterica* subsp. enteric

C;Species: *Salmonella enterica* subsp. enterica serovar Typhi

A;Note: this species has also been called *Salmonella typhi*

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AD0623

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farra

. S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* ser

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AD0623

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-562 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD05454.1; PID:G16502215; GSPDB:GN00176

C;Genetics:

A;Gene: STV1061

Query Match 34.6%; Score 47; DB 2; Length 662;

Best Local Similarity 40.9%; Pred. No. 67;

Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 3 VFTSTEMANKAAEAVLKGQVETI 24

Db 600 IYSELINKAAVAGISGKTEV 621

RESULT 12

T29695

hypothetical protein T18H9.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani  
A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtili*  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: C69962  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-363 <GUN>  
A;Cross-references: GB:Z99116; GB:AL009126; NID:G2634723; PIDN:CAB14338.1; PID:G26348  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: ygiU  
C;Superfamily: acetate kinase

Query Match 33.8%; Score 46; DB 2; Length 363;  
Best Local Similarity 71.4%; Pred. No. 50;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 12 AAEAVLKGVETIV 25  
||| ||| ||| |||  
Db 293 AASALKGEVAIV 306

RESULT 15  
E95887  
Probable aminotransferase protein [imported] - *Sinorhizobium meliloti* (strain 1021) m  
C;Species: *Sinorhizobium meliloti*  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: E95887  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing e  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: E95887  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-461 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC48765.1; PID:g15140238; GSPDB:GN00167  
R;Experimental source: strain 1021, megaplasmid pSymB  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
A>Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A;Reference number: A96039; MUID:21368234; PMID:11474104

A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A;Reference number: A96039, UID:21368234, PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: Smb20379  
A;Genome: plasmid

Query Match 33.8%; Score 46; DB 2; Length 461;  
Best Local Similarity 41.7%; Pred. No. 65;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 4 FSTEMANKAAEAVLKGQVETIVSF 27  
:|||||:|:|:|:|:  
Db 199 FATRMANNLEAILLEGPEETAAAF 222

Search completed: November 13, 2003, 09:29:14  
Job time : 6.81323 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 2.50584 Seconds  
(without alignments)  
525.472 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136

Sequence: 1 VYVFSTEMANKAAEAVLKGQVETIVSPH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	1426	1 BCL9_HUMAN	Q00512 homo sapien
2	98	72.1	1469	1 BCL9_DROME	Q961d9 drosophila
3	60	44.1	1205	1 HEM2_CLOJO	Q59495 clostridium
4	51	37.5	1217	1 SYV_FUGRU	P49696 fugu rubrip
5	46	33.8	330	1 QUTG_EMENI	P25416 emeritella
6	46	33.8	363	1 BUK_BACSU	P54532 bacillus su
7	46	33.8	2273	1 HFAL_YEAST	P32874 saccharomyc
8	45.5	33.5	224	1 SODM_CHAFE	Q96347 charybdis f
9	45.5	33.5	444	1 GUNN_ERWCA	Q59394 erwinia car
10	45.5	33.5	504	1 GUNV_ERWCA	Q59395 erwinia car
11	45.5	33.5	505	1 GUNV_ERWCA	Q59395 erwinia car
12	45.5	33.5	745	1 ZP2_MACRA	Q77726 macaca radi
13	45	33.1	490	1 PEM3_ARATH	Q9c6b9 arabidopsis
14	44.5	32.7	170	1 AIG2_ARATH	P54121 arabidopsis
15	44	32.4	318	1 YB64_METJA	Q58564 methanococc
16	44	32.4	340	1 MPPC_CAEEL	P40814 caenorhabdi
17	44	32.4	484	1 SGAT_ECOLI	P39401 escherichia
18	44	32.4	602	1 DNUI_AERPE	Q9y418 aeropyrum p
19	44	32.4	629	1 SYV_PYRFU	Q8u149 pyrococcus
20	44	32.4	980	1 SYV_SCHPO	Q75005 schizosacch
21	44	32.4	4427	1 PKSL_BACSU	Q05470 bacillus su
22	43	31.6	119	1 TH12_CAEEL	Q17424 caenorhabdi
23	43	31.6	159	1 R1B4_SCHPO	Q9uub1 schizosacch
24	43	31.6	291	1 EXS3_MYCPN	P75403 mycoplasma
25	43	31.6	345	1 TRPD_SULTO	Q97127 sulfolobus
26	43	31.6	537	1 AREH_SCHPO	Q10269 schizosacch
27	43	31.6	575	1 YDGE_SCHPO	Q10499 schizosacch
28	43	31.6	594	1 SYV_FAT	Q04462 rattus norv
29	43	31.6	602	1 GLSL_HUMAN	Q9u132 homo sapien
30	43	31.6	609	1 NPRO_VIBPR	Q00971 vibrio prot
31	43	31.6	611	1 EMPA_VIBAN	P43147 vibrio angu
32	43	31.6	662	1 VGLF_CDVO	P12569 canine dist
33	43	31.6	895	1 SECA_CVACA	Q19911 cyanidium c

#### ALIGNMENTS

##### RESULT 1

BCL9\_HUMAN  
ID BCL9\_HUMAN STANDARD; PRT; 1426 AA.  
AC Q00512;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 15-SEP-2003 (Rel. 42, Last annotation update)  
DE B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).  
GN BCL9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Petal brain;  
RX MEDLINE=98158621; PubMed=9490669;  
RA Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,  
RA Jadayel D.W., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,  
RA Dyer M.J.S.;  
RT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel  
RT gene (BCL9) at chromosome 1q23";  
RA Blood 91:1873-1881(1998).  
RN [2]  
RP FUNCTION.  
RX MEDLINE=21952490; PubMed=11955446;  
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
RA Murone M., Zuehlbig S., Basler K.;  
RA "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
RT cytopus-to-the nuclear beta-catenin-TCF complex.";  
RL Cell 109:47-60(2002).  
CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.  
CC -!- SUBUNIT: Binds to beta-catenin (CTNBN1), PYGO1 and PYGO2.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,  
CC testis, ovary and small intestine, and at lower levels in spleen,  
CC colon and blood.  
CC -!- DISEASES: Involved in a t(1;14)(q21;q32) chromosomal translocation  
CC found in a patient with precursor B-cell acute lymphoblastic  
CC leukemia (ALL). This translocation leaves the coding region  
CC intact, but may have pathogenic effects due to alterations in the  
CC expression level of BCL9. Several cases of translocations within  
CC the 3' untranslated region of BCL9 have been found in B-cell  
CC malignancies.  
CC -!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.  
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 1391.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; Y13620; CAA73942.1; ALT\_FRAME.  
 DR Genew; HGNC:1008; BCL9.  
 DR MIM; 602597; .  
 DR GO; GO:0007048; P:oncogenesis; TAS.  
 KW Nuclear protein; Chromosomal translocation; Proto-oncogene;  
 Wnt signaling pathway.  
 FT DOMAIN 231 1378 PRO-RICH.  
 FT DOMAIN 347 377 CTNNB1-BINDING.  
 FT DOMAIN 331 335 POLY-PRO 1.  
 FT DOMAIN 514 517 POLY-PRO 2.  
 FT DOMAIN 900 903 POLY-ALA.  
 FT DOMAIN 970 973 POLY-PRO 3.  
 SQ SEQUENCE 1426 AA; 149314 MW; A240A48771687F1B CRC64;

Query Match 100.0%; Score 136; DB 1; Length 1426;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-12;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28  
 DB 177 VYVFSTEMANKAAEAVLKQVETIVSFH 204

RESULT 2  
 BCL9 DROME STANDARD; PRT; 1469 AA.  
 ID BCL9 DROME  
 AC Q961D9; Q9V4D2; PRT; 1469 AA.  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE BCL-9 homolog (legless protein).  
 GN BCL9 OR LGS OR CG2041.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins K.A., Galie R.P.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Rosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush C.D., Kraft G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun Z.,  
 RA Svirskaas R., Tector R., Turner E., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou S., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.; cDNA resource.";  
 RT "A Drosophila full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RN [3]  
 RP SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND  
 IDE-537.  
 RX MEDLINE=21952490; PubMed=11955446;  
 RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
 RA Murone M., Zuelig S., Basler K.; BCL9/legless-mediated recruitment of  
 RT pygopus to the nuclear beta-catenin-TCF complex.";  
 RL Cell 109:47-60(2002).  
 CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.  
 CC -!- SUBUNIT: Binds to ARM and PYGO.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
 CC throughout development.  
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AE003844; NAF59345.1; ALT\_SEQ.  
 DR EMBL; AY051851; AK93075.1; .  
 DR EMBL; AF457205; AAL91368.1; .  
 DR FlyBase; FBgn0039907; lgs.  
 DR GO; GO:0005634; C:nucleus; IEP.  
 DR GO; GO:0030528; F:transcription regulator activity; IPI.  
 DR GO; GO:0010177; P:positive regulation of wnt receptor signal. . . IPI.  
 DR GO; GO:0007367; P:segment polarity determination; IMP.  
 KW Nuclear protein; Developmental protein; Segmentation polarity protein;  
 Wnt signaling pathway.  
 FT DOMAIN 511 555 ARM-BINDING.  
 FT DOMAIN 1134 1173 ASN-RICH.  
 FT DOMAIN 1340 1449 GDN-RICH.  
 FT DOMAIN 1162 1169 POLY-ASN.  
 FT MUTAGEN 514 514 G->E: IN ALLELE LGS-21L.  
 FT MUTAGEN 534 534 L->F: IN ALLELE LGS-17E; SEGMENT POLARITY  
 FT MUTAGEN 537 537 PHENOTYPE.  
 FT MUTAGEN I->K: IN ALLELE LGS-17F.  
 SQ SEQUENCE 1469 AA; 153759 MW; 5672E01E7200ED08 CRC64;

Query Match 72.1%; Score 98; DB 1; Length 1469;  
 Best Local Similarity 57.1%; Pred. No. 1.6e-06;  
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28  
 DB 323 IFVFSTQLANKGAESVLGQFQTITAYH 350

RESULT 3  
 HEM2 CLOJO STANDARD; PRT; 205 AA.  
 ID HEM2 CLOJO  
 AC Q59295;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen synthase) (ALAD) (ALADH) (Fragment).  
 GN HEMB.  
 OS Clostridium josui.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OK NCBI\_TaxID=1499;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN:PERM P-9684;  
 RX MEDLINE=95394829; PubMed=7665501;  
 RA Fujino E., Fujino T., Karita S., Sakka K., Ohmiya K.;  
 RT "Cloning and sequencing of some genes responsible for porphyrin biosynthesis from the anaerobic bacterium Clostridium josui.";  
 RL J. Bacteriol. 177:5169-5175(1995).  
 CC -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate = porphobilinogen + 2 H<sub>2</sub>O.  
 CC -!- COFACTOR: Zinc (By similarity).  
 CC -!- PATHWAY: Siroheme biosynthesis.  
 CC -!- SUBUNIT: Homooctamer (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ALADH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D28503; BAA05863.1; -  
 DR PIR; I40812; I40812.  
 DR HSP; P15002; I84E.  
 DR InterPro; IPR001731; Alad\_dehydratase.  
 DR Pfam; PF00490; ALAD; 1.  
 DR PRINTS; PR00144; DALDHORTASE.  
 DR ProDom; PD002304; Alad\_dehydratase; 1.  
 DR PROSITE; PS00169; D\_ALA\_DEHYDRATASE; PARTIAL.  
 KW Porphyrin biosynthesis; Lyase; Zinc.  
 FT DOMAIN 114 132  
 FT NON TER 205 205  
 FT ZINC-BINDING (BY SIMILARITY).  
 SQ SEQUENCE 205 AA; 23172 MW; 886F9DAEFD81144E CRC64;  
 Query Match 44.1%; Score 60; DB 1; Length 205;  
 Best Local Similarity 46.2%; Pred. No. 0.1;  
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 QY 2 YVFSTEMANKAAEAVLKGOVETIVSF 27  
 Db 51 YHFSPDMVGVKAIENALKADVKSVLLF 76  
 RESULT 4  
 SVF\_FUGRU  
 ID SVF\_FUGRU STANDARD; PRT; 1217 AA.  
 AC P49696;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine-tRNA ligase) (VALRS).  
 GN VAR31.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Takifugu.  
 OK NCBI\_TaxID=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97396021; PubMed=9254008;

RA Lim E.H., Corrochano L.M., Elgar G., Brenner S.;  
 RT "Genomic structure and sequence analysis of the valyl-tRNA synthetase gene of the Japanese pufferfish, *Fugu rubripes*.";  
 RL DNA Seq. 7:141-151(1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate + L-valyl-tRNA(Val).  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC -!- SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION FACTOR 1-GAMMA.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X91856; CAA62967.1; -  
 DR HSP; P96142; IGAX.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR002300; tRNA-synt\_1a.  
 DR InterPro; IPR001412; tRNA-synt\_1.  
 DR InterPro; IPR002303; tRNA-synt\_val.  
 DR Pfam; PF00043; GST\_C; 1.  
 DR Pfam; PF00133; tRNA-synt\_1; 1.  
 DR PRINTS; PR00986; TRNASYNTHVAL.  
 DR TIGRFAMs; TIGR00422; vals; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
 FT DOMAIN 1 320  
 FT SITE 293 303  
 FT SITE 809 813  
 FT BINDING 812 812  
 FT ATP (BY SIMILARITY).  
 SQ SEQUENCE 1217 AA; 138218 MW; 5E08AF24B5C8A7A1 CRC64;  
 Query Match 37.5%; Score 51; DB 1; Length 1217;  
 Best Local Similarity 37.0%; Pred. No. 14;  
 Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;  
 QY 2 YVFSTEMANKAAEAVLKGOVETIVSF 28  
 Db 626 YVSCDMGKQADAVREGRLKLIIPDHH 652  
 RESULT 5  
 QUTG\_EMENI  
 ID QUTG\_EMENI STANDARD; PRT; 330 AA.  
 AC P25416;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE QUTG protein.  
 GN QUTG.  
 OS *Emmericella nidulans* (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; *Emmericella*.  
 OK NCBI\_TaxID=162425;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89181521; PubMed=2978880;  
 RA Hawkins A.R., Lamb H.K., Smith M., Keyte J.W., Roberts C.F.;  
 RT "Molecular organisation of the quinic acid utilization (QUT) gene cluster in *Aspergillus nidulans*.";  
 RL Mol. Gen. Genet. 214:224-231(1988).  
 RN [2]  
 RP SIMILARITY TO INOSITOL MONOPHOSPHATASE.  
 RX MEDLINE=91080861; PubMed=2175387;  
 RA Lamb H.K., Hawkins A.R., Smith M., Harvey I.J., Brown J., Turner G., Roberts C.F.;  
 RT "Spatial and biological characterisation of the complete quinic acid utilisation gene cluster in *Aspergillus nidulans*.";  
 RL Mol. Gen. Genet. 223:17-23(1990).

```

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presser E., Fujic C., Ruchelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrif P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tasaku V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zurestein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL CL
CC -|- CATALYTIC ACTIVITY: ATP + 2-butanolate = ADP + butanoyl phosphate.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC
CC -|- SIMILARITY: Belongs to the acetatekinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84432; BAAL2596.1; -.
DR EMBL; Z99116; CAB14338.1; -.
DR FIR; C69962; C69962.
DR Subtilist; BG11724; buk.
DR HAMAP; MF_00542; -.
DR InterPro; IPR000890; Acetate_kin.
DR Pfam; PF00871; Acetate_kinase; 1.
DR PRINTS; PR00471; ACETATEKINASE.
DR PROSITE; PS01075; ACETATE_KINASE 1; 1.
DR PROSITE; PS01076; ACETATE_KINASE 2; 1.
KW Transferase; Kinase; Complete proteome.
SQ SEQUENCE 363 AA; 39764 MW; 94ADAS1211F98DA0 CRC64;
Query Match 33.8%; Score 46; DB 1; Length 363;
Best Local Similarity 71.4%; Pred. NO. 23;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 12 AREAVLKGVETIV 25
DB 293 ASAAALKGVEAIV 306
|||||
|||
RESULT 7
HFAL YEAST STANDARD; PRT; 2273 AA.
ID ID HFAL YEAST
AC P32874;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DE HFAL protein.
GN HFAL OR YMR207C OR YM8261.O1C OR YM8325.O8C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX RX
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentiles S., Hamlin N., Hunt S.,
RA Jagella K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";

```





```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39788; AAC37033.1; -.
CC HSSP; O85465; 1A3H.
CC InterPro; IPR001956; CBD 3.
CC Pfam; PF00942; CBM 3; 1.
CC ProDom; PD001947; CBD 3; 1.
CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC SIGNAL 1 31
CC CHAIN 32 444
CC DOMAIN 32 444
CC ACT_SITE 168 168
CC ACT_SITE 256 256
CC SEQUENCE 444 AA; 48300 MW; FA784179004CB943 CRC64;
CC -----
CC Query Match 33.5%; Score 45.5; DB 1; Length 444;
CC Best Local Similarity 36.7%; Pred. No. 33;
CC Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;
CC -----
CC QY 2 YVFSTEMANKAEAVLKGO---VETIVSFH 28
CC Db 101 YISNPSLANKVKEAVAAQSLGVYIIDWH 130
CC -----
CC RESULT 10
CC GUNW_ERWCA STANDARD; PRT; 504 AA.
CC ID GUNW_ERWCA STANDARD; PRT; 504 AA.
CC AC Q59395;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Endoglucanase VI precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
CC DE (Cellulase VI).
CC GN CELV1.
CC OS Erwinia carotovora.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Pectobacterium.
CC OX NCBI_TaxID=554;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=SCC3193;
CC RX MEDLINE=95231512; PubMed=7715600;
CC RA Mae A., Heikinheimo R., Palva E.T.;
CC RT "Structure and regulation of the Erwinia carotovora subspecies
CC RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
CC RT phytopathogenicity.";
CC RL Mol. Gen. Genet. 247:17-26(1995).
CC CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC CC -|- SUBCELLULAR LOCATION: Secreted.
CC CC -|- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79241; CAA55823.1; -.
CC PIR; S54744; S54744.
```

```
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
DR SIGNAL 1 31
DR CHAIN 32 504
DR DOMAIN 32 504
DR ACT_SITE 168 168
DR ACT_SITE 256 256
DR SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;
DR -----
DR Query Match 33.5%; Score 45.5; DB 1; Length 504;
DR Best Local Similarity 36.7%; Pred. No. 38;
DR Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;
DR -----
DR QY 2 YVFSTEMANKAEAVLKGO---VETIVSFH 28
DR Db 101 YIANPSLANKVKEAVAAQSLGVYIIDWH 130
DR -----
DR RESULT 11
DR GUNW_ERWCA STANDARD; PRT; 505 AA.
DR ID GUNW_ERWCA STANDARD; PRT; 505 AA.
DR AC Q47056;
DR DT 01-NOV-1997 (Rel. 35, Created)
DR DT 01-NOV-1997 (Rel. 35, Last sequence update)
DR DT 15-SEP-2003 (Rel. 42, Last annotation update)
DR DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DR DE (Cellulase V).
DR GN CELV.
DR OS Erwinia carotovora.
DR OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
DR OC Enterobacteriaceae; Pectobacterium.
DR OX NCBI_TaxID=554;
DR RN [1]
DR RP SEQUENCE FROM N.A.
DR RC STRAIN=SCR1193;
DR RX MEDLINE=94067016; PubMed=8246888;
DR RA Cooper V.J.C., Salmund G.P.C.;
DR RT "Molecular analysis of the major cellulase (CelV) of Erwinia
DR RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
DR RT domains.";
DR RL Mol. Gen. Genet. 241:341-350(1993).
DR CC -|- FUNCTION: Endoglucanase with some exoglucanase activity.
DR CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
DR CC linkages in cellulose, lichenin and cereal beta-D-glucans.
DR CC -|- SUBCELLULAR LOCATION: Secreted.
DR CC -|- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature
DR CC optimum about 42 degrees Celsius.
DR CC -|- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
DR CC HYDROLASES).
DR -----
DR This SWISS-PROT entry is copyright. It is produced through a collaboration
DR between the Swiss Institute of Bioinformatics and the EMBL outstation -
DR the European Bioinformatics Institute. There are no restrictions on its
DR use by non-profit institutions as long as its content is in no way
DR modified and this statement is not removed. Usage by and for commercial
DR entities requires a license agreement (See http://www.isb-sib.ch/announce/
DR or send an email to license@isb-sib.ch).
DR -----
DR EMBL; X76000; CAA53592.1; -.
DR PIR; S39962; S39962.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR Pfam; PF00942; CBM 3; 1.
DR SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;
```

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 9.15175 Seconds  
(without alignments)  
789,518 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136

Sequence: 1 VVVFSTEMANKAEAVLKGQVETIVSPH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rviro.\*
- 16: sp\_bacterioph.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	55	40.4	1050	5 Q9ULQ4	Q9ULQ4 caenorhabdi
2	51	37.5	243	16 Q97JA0	Q97JA0 clostridium
3	51	37.5	1049	5 Q9V6L1	Q9V6L1 drosophila
4	51	37.5	1049	5 Q960E6	Q960E6 drosophila
5	50	36.8	459	10 Q82521	Q82521 capsicum ch
6	50	36.8	642	4 Q9H6R2	Q9H6R2 homo sapien
7	50	36.8	642	4 Q96GN2	Q96GN2 homo sapien
8	50	36.8	1098	4 Q86002	Q86002 homo sapien
9	49.5	36.4	350	2 Q8GGI9	Q8GGI9 lactobacill
10	49	36.0	141	10 Q9LFW7	Q9LFW7 arabidopsis
11	49	36.0	143	17 Q976C9	Q976C9 sulfolobus
12	49	36.0	187	16 Q8G3S3	Q8G3S3 bifidobacte
13	49	36.0	210	10 Q9FZL1	Q9FZL1 arabidopsis
14	49	36.0	233	17 Q96YA9	Q96YA9 sulfolobus
15	49	36.0	237	17 Q96X55	Q96X55 sulfolobus
16	49	36.0	268	16 Q8DFX2	Q8DFX2 vibrio vuln

17	49	36.0	458	16 Q83532	Q83532 treponema p
18	49	36.0	586	16 Q9KMS4	Q9KMS4 vibrio chol
19	49	36.0	1047	3 Q8TFZ1	Q8TFZ1 aspergillus
20	48.5	35.7	311	16 Q9FIC9	Q9FIC9 campylobact
21	48.5	35.7	688	5 Q96497	Q96497 trypanosoma
22	48	35.3	225	17 Q96250	Q96250 sulfolobus
23	48	35.3	235	17 Q976P7	Q976P7 sulfolobus
24	48	35.3	330	16 Q9K7G4	Q9K7G4 bacillus ha
25	48	35.3	609	10 Q9LV35	Q9LV35 arabidopsis
26	48	35.3	639	11 Q8BIN9	Q8BIN9 mus musculu
27	48	35.3	1034	3 Q9Y7E9	Q9Y7E9 candida gla
28	48	35.3	1659	4 Q81WQ9	Q81WQ9 homo sapien
29	47.5	34.9	444	16 Q8XKX5	Q8XKX5 clostridium
30	47	34.6	161	2 Q8GJ63	Q8GJ63 bartonella
31	47	34.6	319	10 Q81460	Q81460 arabidopsis
32	47	34.6	352	17 Q980E9	Q980E9 sulfolobus
33	47	34.6	399	16 Q8P1N0	Q8P1N0 streptococc
34	47	34.6	476	10 Q8RVV4	Q8RVV4 lycopersico
35	47	34.6	581	11 Q922Q3	Q922Q3 mus musculu
36	47	34.6	662	16 Q827U3	Q827U3 salmoneila
37	47	34.6	732	16 Q8EA14	Q8EA14 shewanella
38	47	34.6	1072	3 Q8WZK5	Q8WZK5 neurospora
39	47	34.6	1261	5 Q8WMT3	Q8WMT3 dictyosteli
40	47	34.6	2388	5 Q8MYG5	Q8MYG5 dictyosteli
41	47	34.6	2473	12 Q71209	Q71209 grapevine 1
42	46.5	34.2	541	16 Q8EFD3	Q8EFD3 shewanella
43	46.5	34.2	614	5 Q22551	Q22551 caenorhabdi
44	46.5	34.2	771	5 Q19380	Q19380 caenorhabdi
45	46	33.8	320	5 Q81JX6	Q81JX6 plasmodium

## ALIGNMENTS

RESULT 1  
Q9ULQ4 PRELIMINARY; PRT; 1050 AA.  
ID Q9ULQ4;  
AC Q9ULQ4; (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Y87G2A.5 protein.  
GN Y87G2A.5  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA White S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL; AL110500; CAB60428.1; -.  
DR HSRP; P96142; IGAX.  
DR WormPep; Y87G2A.5; CE24685.  
DR InterPro; IPR002300; tRNA-synt\_1a.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR InterPro; IPR002303; tRNA-synt\_val.  
DR Pfam; PF00133; tRNA-synt\_1; 1.  
DR PRINTS; PR00986; TRNASYNTHVAL.  
DR TIGRFBMS; TIGR00422; vals; 1.  
DR PROSITE; PS00178; AA TRNA\_LIGASE\_I; 1.  
SQ SEQUENCE 1050 AA; 118920 MW; F33DBS3587EAC057 CRC64;

Query Match 40.4%; Score 55; DB 5; Length 1050;  
Best Local Similarity 44.4%; Pred. No. 22; Indels 0; Caps 0;  
Matches 12; Conservative 2; Mismatches 13;

[illegible]

DR PROSITE; PS00178; AA TRNA LIGASE I; 1;  
SQ SEQUENCE 1049 AA; 118253 MW; 13ASU3ABF69E9EBB CRC64;

Query Match 37.5%; Score 51; DB 5; Length 1049;  
Best Local Similarity 40.7%; Pred. No. 89;  
Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVALKQGVETIVSPH 28  
DB 471 YVSCSDMAASATEAVRSGELKIPEHH 497

## RESULT 4

Q960E6 PRELIMINARY; PRT; 1049 AA.  
ID Q960E6  
AC Q960E6  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE SD04748P.  
GN AATS-VAL OR CG4062  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda C.J.,  
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RE EMBL; AY052099; AK93523.1;  
DR FlyBase; FBgn027079; Aats-val.  
DR InterPro; IPR002300; tRNA-synt\_1a.  
DR InterPro; IPR01412; tRNA-synt\_1.  
DR InterPro; IPR002303; tRNA-synt\_val.  
DR Pfam; PF00133; tRNA-synt\_1; 1.  
DR PRINTS; PR00986; tRNA-synt\_val.  
DR TIGRFAMs; TIGR00422; vals; 1.  
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.  
SQ SEQUENCE 1049 AA; 118331 MW; 56F322C7414BEAC4 CRC64;

Query Match 37.5%; Score 51; DB 5; Length 1049;  
Best Local Similarity 40.7%; Pred. No. 89;  
Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVALKQGVETIVSPH 28  
DB 471 YVSCSDMAASATEAVRSGELKIPEHH 497

## RESULT 5

Q82521 PRELIMINARY; PRT; 459 AA.  
ID Q82521  
AC Q82521  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative aminotransferase.  
OS Capsicum chinense.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiales; Solanales; Solanaceae; Capsicum.  
OX NCBI\_TaxID=80379;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=habanero;  
RA Aluru M., Curry J., O'Connell M.;  
RT "Nucleotide sequence of a Probable Aminotransferase Gene (Accession  
No. AF085149) from Habanero Chile. (FGR98-182).";

RL Plant Physiol. 118:1102-1102(1998).  
DR EMBL; AF085149; AAC78480.1; -;  
DR HSP; P04181; 20AT.  
DR InterPro; IPR005814; Aminotrans\_3.  
DR Pfam; PF00202; aminotrans\_3; 1.  
DR PROSITE; PS00600; AA TRANSFER\_CLASS\_3; 1.  
KW Aminotransferase; Transferase.  
SQ SEQUENCE 459 AA; 50729 MW; 02ABB4D728B524E4 CRC64;

Query Match 36.8%; Score 50; DB 10; Length 459;  
Best Local Similarity 41.7%; Pred. No. 51;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVALKQGVETIVSF 27  
DB 196 FSTRANLESLLKSGPEIVAAF 219

## RESULT 6

Q9H6R2 PRELIMINARY; PRT; 642 AA.  
ID Q9H6R2  
AC Q9H6R2  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein FLJ21965.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hixji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RE "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK025618; BAB15191.1; -;  
DR HSP; P96142; IGAX.  
DR InterPro; IPR002300; tRNA-synt\_1a.  
DR InterPro; IPR002303; tRNA-synt\_val.  
DR Pfam; PF00133; tRNA-synt\_1; 1.  
DR PRINTS; PR00986; tRNA-synt\_val.  
KW Hypothetical protein.  
SQ SEQUENCE 642 AA; 71578 MW; C9E37EE1D74287F1 CRC64;

Query Match 36.8%; Score 50; DB 4; Length 642;  
Best Local Similarity 44.4%; Pred. No. 73;  
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVALKQGVETIVSPH 28  
DB 60 FVRCQEMGARAARAKAVESGADELSPSPH 86

## RESULT 7

Q96GN2 PRELIMINARY; PRT; 642 AA.  
ID Q96GN2  
AC Q96GN2  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RE EMBL; BC009355; AAN09355.1; -;

RP SEQUENCE FROM N.A.  
RC STRAIN=CCM3626;  
RA Brinzel F., Hubert J.-C.;  
RT "Investigation of arginine requirement in Lactobacillus isolated from  
RT different environments revealed point mutations, insertions and  
RT deletions in arginine biosynthetic genes."; ;  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF514870; AA015990.1; -.  
SQ SEQUENCE 350 AA; 38178 MW; 64D1986ED73C8AAA CRC64;

Query Match 36.4%; Score 49.5; DB 2; Length 350;  
Best Local Similarity 54.5%; Pred. No. 45;  
Matches 12; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 3 VFSTEWANKAAEAVALKGOVETI 24  
|::: |||| |  
Db 176 VYSTDLLAKAE--KGQVDAL 194

RESULT 10  
Q9LEFW PRELIMINARY; PRT; 141 AA.

ID Q9LEFW AC Q9LEFW;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE T7N9.31.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
[1]  
SEQUENCE FROM N.A.  
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RA "genomic sequence for Arabidopsis thaliana BAC T7N9 from chromosome  
RA i.";  
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.;  
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
[6]  
SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Theologis A., Ecker J.;  
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

```

RA Primrose R.D., Arigoni F.:
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AE014802; AAN25468.1;
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 187 AA; 20617 MW; E697C3C127277D81 CRC64;

Query Watch 36.08; Score 49; DB 16; Length 187;
Best Local Similarity 45.2%; Pred. NO. 26;
Matches 14; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

QY 7 EMANKAAE-----AVLKQGVETIVSF 27
      ||| ||| ||| ||| ||| |||
DB 26 EWAALASEDYEDKNPLLVALKGAVNTLVAF 56

RESULT 13
Q9FZL1
ID Q9FZL1 PRELIMINARY; PRT; 210 AA.
AC Q9FZL1;
DT 01-NAR-2001 (TrEMBLrel. 16, Created)
DT DT 01-MAR-2001 (TrEMBLrel. 15, Last sequence update)
DT DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE F17121.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,
RA Shinn P., Alcafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Tokumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RA "Genomic sequence for Arabidopsis thaliana BAC F17121 from chromosome
RA 1.";
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Alcafi H., Bei B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Tokumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004557; AAF99722.1; -.
DR DR InterPro; IPR003822; PAH.
DR PFam; PF02671; PAH; 1.
SQ SEQUENCE 210 AA; 24697 MW; FCD8130CD75700A0 CRC64;

```



Query Match 36.0%; Score 49; DB 10; Length 210;  
Best Local Similarity 41.7%; Pred. No. 30;  
Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSF 27  
DB 181 WSPFSTNKAADRLKGELENNVTF 204

## RESULT 14

ID Q96YA9 PRELIMINARY; PRT; 233 AA.  
AC Q96YA9  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical protein ST2259.  
GN ST2259.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AF000989; BAB67368.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 233 AA; 27253 MW; 8B9B6FDB60EE36D CRC64;

Query Match 36.0%; Score 49; DB 17; Length 233;  
Best Local Similarity 39.3%; Pred. No. 34;  
Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 VVVFSTEMANKAAEAVLKQGVETIVSFH 28  
DB 120 VLVVLTNNMKNVEKAEKVLKTRIDKVVYIH 149

## RESULT 15

ID Q96X55 PRELIMINARY; PRT; 237 AA.  
AC Q96X55  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein ST1165.  
GN ST1165 OR ST0857.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";

RL DNA Res. 8:123-140(2001).  
DR EMBL; AF000985; BAB66201.1; --  
DR EMBL; AF000984; BAB65870.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 237 AA; 27691 MW; B931759A9980D362 CRC64;

Query Match 36.0%; Score 49; DB 17; Length 237;  
Best Local Similarity 43.3%; Pred. No. 34;  
Matches 13; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 VVVFSTEMAN--KAAEAVLKQGVETIVSFH 28  
DB 120 VLVVLTNNMKNVEKAEKVLKTRIDKVVYIH 149

Search completed: November 13, 2003, 09:27:36  
Job time : 12.1518 secs

